

The cloud over Sarawak

Forest fires and urban pollution are wreaking havoc in South-East Asia. Technology and ecology can offer a little help, but more national and international awareness would also be useful.

People in the West could be excused for having missed it, but an environmental disaster has been developing in Indonesia, Malaysia and Singapore. Barely noticed by the Western media over the past two weeks, the region has been hit by atmospheric pollution at levels seriously affecting millions of people, with hundreds of deaths because of consequently impeded emergency relief from another calamity — drought. The situation is worsening as we go to press.

The calamity highlights two severe problems. One is potentially lethal smog that has arisen in urban areas due to a combination of factors — in that sense, economic progress has made unhealthy places of Kuala Lumpur and Sarawak, where vehicle and industrial pollution regularly produce dense smogs. Add to that the plumes from forest fires now raging in upwind Indonesia and it is small wonder that world pollution records are being broken (see page 321). The only silver lining on this cloud from Kalimantan and Sumatra is that it could put additional pressure on rapidly developing nations to clean up their acts.

The second problem reflected in these events is that of devastating forest management. This year, the annual burn of forest clearings in Indonesia has spread out of control largely due to the drought — itself due to the intense El Niño. Indonesia is 75 per cent (rapidly diminishing) forest which, whatever may be said to the contrary by Indonesian agencies, is poorly protected and policed in the face of destructive practices and the economic pressures that give rise to them. Indeed, there is every reason to believe that fires on this scale result from clearances partly aided by the government to promote migration, rather than slash-and-burn agriculture.

Forest clearances for the timber industry — 25 per cent of the country's industrial exports — give rise to a species of broad-leaved grass that grows densely and acts like a tinderbox. Once it starts to

burn, there is little anyone can do but pray for heavy rain. But experience in similar terrain in northern Australia, for example, has demonstrated the value of controlled burning: choosing the scale on which to burn, controlling spread by firebreaks and by the appropriate frequency and timing with respect to the weather. Replanting with trees to block the light and thus cut back the grassland can also be desirable.

Obstacles to such approaches in Indonesia are no doubt significant. In the long term, spreading awareness of environmental consequences is probably as substantial a contribution to solving the problem as any other. That awareness can be fostered not only by campaigns by governments but also with technology: the growing availability of remote-sensing images can assist in minimizing damage and in the appreciation of the scale of potential problems.

Attempts at international forestry agreements are constantly undermined by national pressures — as Venezuela, for example, is demonstrating in its blocking of forestry action within the United Nations biodiversity convention. Such agreements are based on the principle that forestry management is primarily a national responsibility, but the international perspective they bring — of best practice and feasible options, for example — can only help the resolve of countries in tackling such problems.

It is all the more regrettable, therefore, that international awareness is low and diminishing in developed countries which themselves contribute to forest devastation. The indifference to South-East Asia currently exhibited by Western media stands in marked contrast to, for example, the epic response to minor health problems among some fishermen on the east coast of the United States (see pages 317–318). Regrettably, that reflects a growing insularity in the West, both in the media and in the populations that they serve and reflect. □

Ashkenazi signals are timely

Concerns in the US Jewish community about the potential misuse of genetic results need to be addressed.

One of the profound ironies of modern science is that no social group is better equipped either in its ability to provide reliable, ordered health data or in its capacity to interpret its genetic and medical implications than the Jewish community. At the same time, no group has greater reason to be wary of the way in which such implications can be misused.

That issue has now prompted some leaders of the US Jewish community to seek reassurance that research data from studies of particular ethnic groups are not used to those groups' disadvantage (see page 322). Their concern is that the publicity surrounding recent research — such as last year's discovery of a significant proportion of a particular breast cancer mutation among Ashkenazi Jews in the United States, and similar recent findings related to colon cancer — may indirectly create problems for members of that group, for example in obtaining adequate health insurance cover.

Several important points need to be remembered. One is that a high incidence of a particular mutation does not necessarily imply a

high incidence of the related disease; other, perhaps environmental or psychological, factors may be equally important. The second is that, for understandable reasons of research methodology, most recent genetic 'discoveries' have been associated with well-structured — and often geographically isolated — social groups, such as the inhabitants of Tristan da Cunha (asthma), Iceland (familial essential tremor) and Finland (diastrophic dysplasia).

Unsurprisingly, the diseases chosen for study in these groups tend to be those with a significantly higher presence than in the rest of the population. But the attractiveness of searching for other genetic clues in these same populations should not blind researchers to the potential risks. The real challenge is eliminating the stigma still attached to the concept of 'mutant'. This is a question partly of education, partly of legislation (for example, banning genetic discrimination in insurance and employment). If the flag raised by the Jewish community helps to speed either process, it will have achieved a worthwhile goal. □



Scientists close in on 'cell from hell' lurking in Chesapeake Bay

[WASHINGTON] A mystery organism which is believed to have killed thousands of fish in Chesapeake Bay, on America's mid-Atlantic coast, and prompted a major public health scare, should become less mysterious within weeks as scientists rush to track down the organism and identify the toxins it releases.

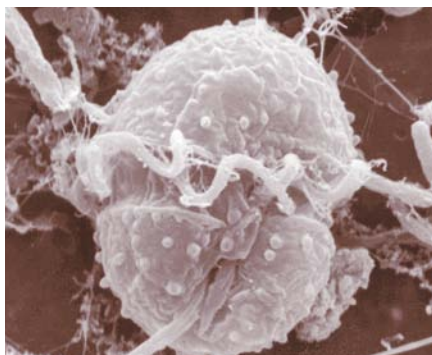
Multi-agency research is already under way to investigate the dinoflagellates, which are thought to have killed some fish and caused large lesions on many others. According to an early study, fishermen who work the infected waters have suffered memory loss and skin problems. The infected fish are menhaden, which is never eaten but is harvested in huge quantities for use in pet-food and as a source of cheap oil for margarine.

But the scientists who pioneered the study of similar organisms claim their investigations have been obstructed by officials in the nearby state of North Carolina, who wanted to protect their tourism, fisheries and agriculture industries from the kind of public attention now developing in Maryland.

The organisms in the Chesapeake — a 200-mile-long sea inlet whose health and conservation is a major public concern in the region around Washington — are closely related to *Pfiesteria piscicida*, a remarkable single-cell organism first identified five years ago by JoAnn Burkholder of North Carolina State University (see *Nature* 358, 407; 1992).

Pfiesteria can take twenty or more distinct forms during its life cycle. It can play dead on the sea bed, grow via photosynthesis, secrete toxins, eat the fish poisoned by the toxins, and return to an ambient state — to be branded "the cell from hell" by *The New York Times*.

"Of all the organisms I know, it's the most complex by a quantum jump," says Donald Anderson of the Woods Hole Oceanic Institute, Massachusetts. "It's hard to understand how it evolved that way. But if you wanted an



Toxic shocker: the organisms blamed for fish kills are related to *Pfiesteria piscicida* (left), whose discoverer, JoAnn Burkholder (right), blames it for neurological symptoms she has experienced.



organism that would survive for millions of years, that would be a good way to do it."

Burkholder found that *Pfiesteria* growth was boosted by the presence of phosphorus in the water, and became embroiled in arguments with North Carolina officials about the role of the effluent from the state's hog-farming industry in the organism's behaviour. Burkholder and Howard Glasgow, her collaborator, suffered neurological symptoms which they attributed to their contact with *Pfiesteria* in the laboratory. Their story is related in a recently published melodrama*.

Ironically, the suspicion that *Pfiesteria* is a health hazard has slowed down research on the organism. Scientists could not study it unless their laboratories met standards for the containment of biological hazards, which are rare in marine biology laboratories.

But since fish kills were reported in Chesapeake Bay just over a month ago, Burkholder and a few others have won access to medical laboratories for the analysis, and have made substantial progress in identifying the toxins released by *Pfiesteria*. They have also been adopted as advisers to the state gov-

* *And the Waters Turned to Blood*, by Rodney Barker, \$24 from Simon & Schuster.

ernments confronting the problem.

Burkholder is working with the Medical University of South Carolina to track down a toxin associated with brain damage in fish. Edward Noga, her *Nature* co-author, has worked with researchers at the University of Miami on a separate toxin that may cause lesions on the fish. "We've made great strides in the past three weeks, thanks to the help of these places," Burkholder says. "We've isolated a water-soluble toxin and a lipid-soluble toxin, and are very close to finding out what the water-soluble toxin is."

But Karen Steidinger of the Florida Marine Research Institute, who has been analysing water samples from Chesapeake Bay, says that the two dinoflagellates she has identified are not *Pfiesteria*. "Once you take off their membrane, they are very different," she says. Although closely related to *Pfiesteria*, these organisms could produce different toxins that interact differently with fish and with people, she says.

A study conducted in the past three weeks for the Maryland state government, by a team led by Glenn Morris of the University of Maryland School of Medicine, found a strong correlation between exposure to water near the fish kill and various neurocognitive symptoms. The study compared a self-selected sample of eleven fishermen who worked the Pocomoke river, where a major fish kill occurred in August, with a group which may have received lower exposure, and a third group which works nearby on the Atlantic Ocean and is assumed to have had no exposure at all. "We found a significant problem with memory loss" in the high-exposure group, Morris says. "There do appear to be clear, documentable health effects."

Steidinger, however, says that the correlation itself does not prove that people have been poisoned by the *Pfiesteria*-like organisms. The dead fish themselves, she points out, will put toxins in the water: cause and

EPA to be sued over gene-modified crops

[WASHINGTON] The US Environmental Protection Agency (EPA) learned last week that a coalition of consumers, environmentalists and organic farmers intends to sue it for allowing the planting of crops genetically engineered to produce an insecticidal toxin made by *Bacillus thuringiensis* (Bt) a soil bacterium.

Thirty signatories filed a petition with the EPA demanding that it revoke 11 registrations it has issued to five companies to plant modified corn, cotton and potatoes. The petition also asks the agency to cease

granting such permissions, and to complete a statement analysing the environmental impact of Bt crops approved so far.

The petitioners claim that commercial planting of the crops threatens to create widespread resistance to Bt toxin, which is heavily relied on by organic farmers.

Al Heier, an EPA spokesman, said the agency "will consider the petition very seriously". But he argued that the EPA won approval from a panel of outside scientists before agreeing to register the first Bt crop in 1995.

Meredith Wadman

effect can be established only by identifying and tracking the toxins from the organisms.

The Morris study, which was released last week, has been the subject of some disagreement between Maryland and Virginia, the two states surrounding the bay, with the former unwilling at first to hand over its raw data for scrutiny by the latter.

The Democrat governor of Maryland, Parris Glendening, who is up for re-election next year, has sought to impress the largely urban voters of his state by taking immediate, high-profile action on the fish kills. He has closed several rivers to fishing, and obtained federal assistance and pledges of public concern from President Bill Clinton.

Virginia is taking a far more cautious approach. George Allan, its Republican governor, has refused to close any rivers apart from the Pocomoke itself, despite the discovery of fish with lesions in other waterways.

The two men buried some of their differences at a 'governors' summit' hosted by Glendening last week in Annapolis, Maryland's capital, at which they agreed a five-point plan to cooperate in addressing the problem. Allan continues to express doubts about the Morris study, however. "The number of people who reported memory loss was, I believe, eight," he says. "Our medical professionals haven't had the chance to look at it yet. I'd want our folks to look over it and see if they come to the same conclusions."

A barrage of small-scale federal initiatives was announced last week to pursue the mystery organism. The Environmental Protection Agency (EPA) and the National Oceanic and Atmospheric Administration have sent teams to the bay and will provide Maryland with emergency grants.

The Centers for Disease Control, which is expected to receive an emergency \$7-million appropriation from Congress for *Pfiesteria* research, will host a conference in Atlanta next week to "plan a comprehensive public health approach to *Pfiesteria*".

The Food and Drug Administration is expected to explore any possibility that toxins from the organism could enter the food chain through menhaden. There is no evidence so far that the toxins accumulate in the fish, officials say.

Pressure for action to restrict the run-off of nutrients from Maryland's large chicken industry in response to the incidents appears to be waning, however. A plan is already in place to cut these by 40 per cent, and Glendening says he has no plans to legislate further.

Carol Browner, EPA's administrator, who attended the Annapolis meeting, said that the fish kills were "a clarion call" to address the flow of nutrients and other pollution into natural waterways. Not everyone is convinced, however. As Karen Steidinger points out, Spanish explorers of the Florida coast were warned about annual fish kills by the Indians — in 1560.

Colin Macilwain

Cancer scientist takes on top job at Wellcome Trust

[LONDON] Britain's Wellcome Trust, which describes itself as "the world's biggest medical research charity", has chosen a prominent cancer researcher to lead it into the next century.

Mike Dexter, only recently appointed director of the Paterson Institute in Manchester, which is largely funded by the Cancer Research Campaign, will succeed Dame Bridget Ogilvie on her retirement next year.

There had been discussion about whether the trust, whose capital assets, estimated to be worth more than £10 billion (\$16 billion), are even larger than those of the Howard Hughes Medical Institute in the United States, might turn to the business community to recruit its new chief executive.

Instead, the trustees have chosen a widely respected haematologist with experience both as a researcher — he has published more than 300 scientific papers — and in research administration as, for example, chairman of the Medical Research Council (MRC)'s molecular and cellular medicine board.

As one of the three co-founders of Therxsys, Britain's first gene-therapy start-up company which is partly owned by the MRC and has rights to key MRC research on gene vectors, Dexter also has direct experience of working on the commercial applications of scientific discoveries, an increasingly important dimension of the trust's activity.

Dexter says the trust, which has recently taken a lead in establishing a clear career structure for research scientists, "has a responsibility to stand up for the scientific community, and to make sure that funding goes to where the science is excellent, not



Dexter: believes the trust 'has a responsibility to stand up for the scientific community'.

necessarily where some other organization would like it to go".

Having been closely involved in gene-therapy research in Manchester, Dexter says he is keen to support the trust's growing involvement in issues concerned with the public understanding of science — and particularly exploring the implications of modern genetics. "It is important that we bring the public along with us," he says.

Dexter's appointment is described as "inspirational" by Gordon McVie, director-general of the Cancer Research Campaign, which has supported Dexter's research since the early 1970s.

"He has a wealth of experience in biology and a terrific grasp of the technical side of experiments, and the fact that he comes from the cancer side of things is no drawback, as we have been closely linked with many important developments in genetics in recent years," says McVie. Ironically, Sir Henry Wellcome specified in his will setting up the trust in 1936 that no money was to be used for cancer research as such.

David Dickson

£10 million pledged for UK synchrotron source

[LONDON] In an unprecedented move, the Wellcome Trust has announced it is prepared to contribute £10 million to the costs of a new second-generation synchrotron source, known as Diamond, at the Daresbury laboratories in Cheshire, England.

The offer comes as various research councils have been asked by the Office of Science and Technology to give their views on a replacement for Daresbury's current 2-GeV Synchrotron Radiation Source (SRS) which becomes obsolete early in

the next decade.

The heavy use of the facility by researchers has left little doubt that the research councils will be positive. But that will leave open the question of who is prepared to pay for what.

Current estimates are that a new storage ring will cost £100 million — a sum the research councils hope will be found by central government — and that experimental beam lines would cost another £40 million, which the research councils might find out of their own budgets.

Officials of Wellcome Trust have in the past said that the trust should not be expected to make up for the government's failures to maintain a healthy research infrastructure, so the trust may be reluctant to contribute towards the costs of the ring itself.

Despite this uncertainty, David Norman, director of the SRS facility, says he hopes the Wellcome offer will "catalyse" thinking about the value of the facility both in the research councils and in government circles.

D. D.

US biologists adopt cloning moratorium

[WASHINGTON] One of the leading US professional associations of biologists has adopted a voluntary five-year moratorium on the cloning of human beings. But it is also seeking to keep open a window that would allow research on human embryos that might otherwise fall under a wider ban on cloning-related research.

The Federation of American Societies for Experimental Biology (FASEB) last week announced its approval of the moratorium, which had won unanimous approval on 10 September from the federation's public affairs executive committee.

"We want to reassure Americans that biologists have no intentions of cloning human beings," Ralph Yount, president of FASEB and a chemist at Washington State University in Pullman, said in a statement. "Indeed, we would regard cloning a human being as an unethical and reprehensible act."

FASEB has 14 member societies representing more than 52,000 scientists. The moratorium defines "cloning human beings" as "the duplication of an existing or previously existing human being by transferring the nucleus of a differentiated, somatic cell into an enucleated human oocyte, and implanting the resulting product for intra-uterine gestation and subsequent birth".

But text accompanying the moratorium makes a point of contrasting cloning intended for implantation and for *in vitro* research. "We expect that further research using human cells will also be necessary to secure the benefits of insights from animal cloning and nuclear transfer as applied to human health," it states.

In the accompanying statement, FASEB also warns against "imprecise or misused technical language" in planned federal and state legislation. "Such laws could hinder vital biomedical research," it says.

One bill introduced by Representative Vern Ehlers (Republican, Michigan) was amended in July by the House of Representatives Science committee to ban federal funding for the use of cloning for *in vitro* research in human embryos as well as for producing human beings (see *Nature* 388, 505; 1997).

Although the FASEB moratorium defends "important new research" in human cells made possible by cloning, officials deny they are taking a stand in favour of human embryo research. The FASEB board "wanted to deal only with" the issue of most concern to the public, which is cloning a new adult human being, says Mike Stephens, a lobbyist for the group. The group's position on human embryo research is "still being discussed by FASEB and its societies", he adds.

Some feel that FASEB may be trying to avoid provoking controversy by steering clear of the heated debate on the ethics of embryo

research. "They wanted to reassure the American people without necessarily stirring up other things," says Brigid Hogan, professor of cell biology at Vanderbilt University School of Medicine in Nashville, Tennessee.

Hogan co-chaired a panel on embryo research set up by the National Institutes of Health which concluded in 1994 that such research was acceptable for federal funding if it was carried out within strict limits.

Arthur Caplan, a bioethicist at the University of Pennsylvania, agrees with Hogan. "By omission they are hoping that research on some human clones at the embryo level might be allowed," he said.

FASEB officials also deny that the moratorium is intended to respond to the House Science committee's July vote to outlaw the use of cloning technology for *in vitro* research on human embryos. "It was really not our goal to affect legislation," says William Brinkley, the group's vice-president, who is dean of the Graduate School for Biomedical Sciences at Baylor College of Medicine in Texas.

Roger Pedersen, a member of FASEB's public affairs executive committee and a

reproductive geneticist at the University of California, San Francisco, said that member scientists should be able to use the new cloning technology to do research on human cells *in vitro*. This, he said, would help scientists understand how adult nuclei are reprogrammed by cellular cytoplasm, possibly opening avenues to novel ways of repairing and regenerating human tissues.

Pedersen led an earlier move by the 2,000-member Society of Developmental Biologists (SDB), a FASEB member society, to adopt the moratorium. About a quarter of SDB's members participated in an electronic and mail-in vote on the moratorium in early September; 93 per cent approved it.

In a report to President Bill Clinton in June, the National Bioethics Advisory Commission called on scientific societies to comply voluntarily with a moratorium on federal funding for cloning human beings (see *Nature* 387, 644; 1997).

Clinton made a similar plea when he announced the moratorium in March (see *Nature* 386, 97; 1997), and has also drafted legislation outlawing the cloning of human beings for five years.

Meredith Wadman

End agreed for ozone-destroying pesticide

[MONTREAL] More than a hundred countries have agreed to phase out use of the pesticide methyl bromide, which makes a significant contribution to depletion of the ozone layer.

Meeting last week to strengthen the Montreal Protocol, protocol signatories also agreed to try to reduce smuggling of chlorofluorocarbons (CFCs) by adopting licensing and reviewing compliance procedures. Police and customs officials will be given greater powers to intercept illegal imports and exports of ozone-depleting substances.

Methyl bromide is considered to contribute about 10 per cent to the destruction of the ozone layer. Developed countries are responsible for 80 per cent of its use worldwide.

Developed countries will put an end to using methyl bromide by 2005; developing countries have an extra 10 years. They will also be eligible to use a C\$25 million (US\$18 million) fund to help them convert to alternatives to methyl bromide. Canada has contributed C\$5 million to this fund.

Signatories also agreed to find alternatives to CFCs in medicines that release drugs in controlled doses, such as asthmatic inhalers. They will be expected to begin the transition to non-CFC and methyl bromide alternatives by 1999.

The agreements came on the tenth anniversary of the Montreal Protocol, widely considered the most successful piece of

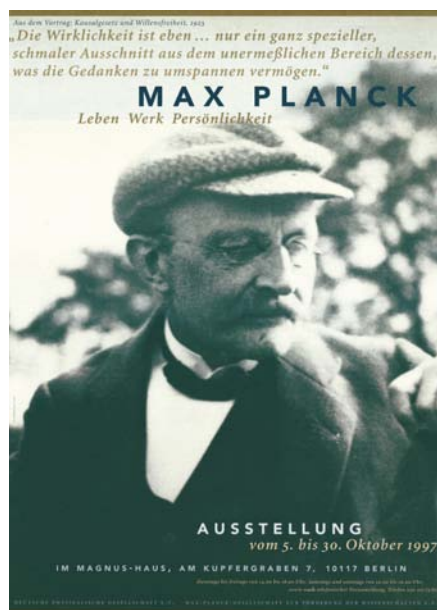
international environmental legislation.

Canada's environment minister, Christine Stewart, welcomed the moves as "a step in the right direction", even though the final outcome fell short of the host country's expectations. Canada had lobbied for earlier deadlines for phasing out methyl bromide — 2001 for developed countries and 2011 for the developing world. But Stewart remained upbeat: "We did not get everything we wanted, but the international community did respond, and this agreement is real progress."

Representatives of industry also welcomed the new agreements. But environmentalist groups were more sceptical, and said the commitments fell short of a total and immediate ban on all sales of ozone-depleting substances in the developed world. Friends of the Earth said: "It's more than a disappointment, it's a tragedy for the planet and the integrity of the Montreal Protocol."

Despite the success of the protocol in reducing ozone-depleting substances, global stratospheric ozone continues to decline. The secretary-general of the World Meteorological Organization, G. O. P. Obasi, told the conference of signatories that "during the past 30 days at the beginning of this year's Antarctic ozone decline, more than 30 per cent of the stratospheric ozone over the southern polar region has been destroyed".

David Spurgeon



German physicists honour Planck

[MUNICH] The fiftieth anniversary of the death of the German physicist Max Planck on 4 October 1947 is to be marked by the Max Planck Society and the German Physical Society with an extensive exhibition describing his life and work.

The exhibition in Magnus House, Berlin, runs from 5 to 30 October. It includes personal items such as Planck's rucksack and icepick — he was a keen mountain hiker — and documents that have not been publicly exhibited before. It will also be the first public showing for his death mask.

More familiar items to be exhibited include the minute books of the German Physical Society which record, for example, the lecture Planck gave to the society in Berlin in 1900 in which he described his law of radiation and introduced for the first time his quantum hypothesis.

Also familiar to Planck aficionados is the 20-minute film commissioned by Joseph Goebbels' ministry of propaganda in 1942, discovered in 1983. The film, which will run continuously throughout the exhibition, features Planck, sitting on a stool, discussing his life. The film was not actually used for propaganda — it was one of a series intended to archive the lives of great Germans.

Magnus House, in former East Berlin, has strong associations with recent German history. The family of Gustav Magnus, the German physicist who in the last century discovered the Magnus effect (the sideways force exerted on a rotating body moving through a fluid), donated the house to the German Physical Society of the former German Democratic Republic in 1958, in honour of the centenary of Planck's birth. After reunification, the house was transferred to the unified German Physical Society. **Alison Abbott**

NASA to fund life science beam line at Brookhaven?

[WASHINGTON] A facility proposed for the Brookhaven National Laboratory in New York state would give space biologists more opportunity to study the effects of cosmic radiation on living organisms — research that is critical if future astronauts are to travel to the Moon and Mars.

The proposed Booster Applications Facility (BAF) could begin construction next year and would probably be funded primarily by the US National Aeronautics and Space Administration (NASA). The White House's science and budget offices have yet to decide how NASA and the Department of Energy, which runs Brookhaven, would split the construction costs for the BAF (about \$20 million), and the \$4 million a year operating costs. Funding would have to be approved by Congress as part of the 1999 budget request.

Richard Setlow, Brookhaven's associate director for life sciences, says that life scientists in NASA have long sought such a facility, as heavy, fast-moving atomic particles pose perhaps the greatest potential health threat to humans venturing beyond Earth orbit.

Astronauts in Earth orbit are protected from galactic cosmic radiation by the Earth's magnetic field. But interplanetary voyagers run the risk of cancer, damage to the nervous system and other problems from bombardment by iron nuclei and other heavy particles.

Little is known about the biological effects of these high-energy particles, says Setlow. The research is hard to do in space because of the difficulty of working with large numbers of test animals and the relative scarcity of radiation in space — the dosages would build up only after many months of exposure. NASA

researchers scrounge about two weeks a year on accelerators at Brookhaven and elsewhere. But that is not enough, Setlow says.

Setlow chaired a National Research Council study last year that called on NASA to boost a programme of ground-based research in this area if it expects to be ready to send humans to Mars in 20 years. But expanded use of existing ground-based facilities would interfere with higher priority physics experiments.

The BAF would divert a new beam line from a circular 'booster' accelerator that shoots pulses of heavy ions into Brookhaven's Alternating Gradient Synchrotron (AGS), which is used by high-energy physicists. Because the AGS receives particles in bursts rather than continuously, the bursts could alternate rapidly between beams without disrupting physics experiments, says Setlow.

Biologists would expose cell cultures and laboratory animals to the radiation. Other scientists would study how different materials perform as radiation shields. NASA may also be interested in the effects of heavy ions on microcircuitry, says Setlow. He expects users for the BAF to number in the hundreds.

There is some urgency to begin construction soon, he says. Diverting a new beam requires breaking into the booster tunnel and installing new steering magnets. This would have to be done before physicists begin using the Relativistic Heavy Ion Collider (RHIC), a major Brookhaven facility due to go on-line in 1999. The collider is fed by the same booster. As a first step, NASA will pay for an environmental impact study of the proposed new facility.

Tony Reichhardt

France asked to drop N-test simulations

[PARIS] Four unions representing French researchers have written a joint letter to Alain Richard, the minister of defence, asking the government to drop its plans for a programme of computer simulations of nuclear weapons explosions, called PALEN.

The unions — the SNCS, the SNTRS-CGT, the SNES-FSU and the SNPEN-FSU — whose members include both university teachers and nuclear physicists, point out that during his election campaign Prime Minister Lionel Jospin wrote that it was "pointless to launch a programme of test simulations that would be costly and unproductive, and could threaten to provoke a reopening of the arms race".

The scientists have sent copies of the letter to President Jacques Chirac, Claude Allègre, the minister of national education,

and Jospin himself.

"We are not seeking unilateral disarmament, merely that France commits to the path opened by the non-proliferation treaty," says Marc Ollivier, the coordinator of the initiative. He points to an interpretation of article 6 of the treaty by the International Court of Justice, stating that "an obligation exists to pursue in good faith up to their conclusion all aspects of negotiations leading to nuclear disarmament".

The unions claim that the PALEN programme does not comply with this commitment, and the researchers ask for the setting up of a broad interdisciplinary programme, involving those engaged in both the natural and the social sciences and associated with the Ministry of Defence, to prepare for disarmament. **Eric Glover**

Forest fires cause pollution crisis in Asia

MIKE HALL/AP

[TOKYO] The Malaysian government last week declared a state of emergency in Sarawak, on the island of Borneo, as smoke pollution from forest fires burning out of control in neighbouring Indonesia enveloped many cities in South-East Asia. In Sarawak, the pollution reached life-threatening levels, forcing schools, factories and offices to shut.

Fires set off by the slash-and-burn activities of plantation owners and farmers in Indonesia are raging out of control in an arc extending around the Malay peninsula from Sumatra through Java to Kalimantan in Borneo and Sulawesi. The fires have been exacerbated by drought thought to have been brought on by what is proving to be one of the most severe El Niño episodes on record.

The situation in Sarawak is particularly severe, as pollution levels have gone off standard scales of measurement, and visibility has been reduced to tens of metres. A state of emergency has been declared, and the 1.9 million people in Sarawak have been advised to stay indoors with windows and doors shut.

Pollution also briefly reached hazardous levels last week in Kuala Lumpur on the west side of the Malay peninsula, and has risen to unhealthy levels in Singapore.

The Indonesian government considered evacuating the city of Rengat in Sumatra, where visibility at one point dropped to a few metres, and people were having difficulty breathing. But these plans have been tem-



Smoky city: air pollution in Kuala Lumpur in Malaysia rose to dangerous concentrations.

porarily abandoned because the situation improved with a change of wind.

Last week, President Suharto of Indonesia apologized to environment ministers from the East Asian region at a meeting in Jakarta, and his government has banned plantation owners and forestry companies from using fire to clear forest.

But a spokesman for the Indonesian government's Environmental Impact and Assessment Agency, quoted in the *Straits Times* of Singapore, said satellite pictures showed that the number of new fires had increased since the ban, as companies rushed

to clear forest before the clampdown.

At least 70 major fires are raging out of control in Kalimantan and central and southern Sumatra. Satellite images obtained by the Centre for Remote Imaging and Sensing (CRISP) of the National University of Singapore show that more than half of the vegetation in southeast Kalimantan has been destroyed by slash-and-burn clearing in the past three months. The Indonesian air force is flying sorties in Kalimantan and Sumatra to seed clouds to induce rain to help clear the pollution.

Datuk Chua Jui Meng, the Malaysian minister of health, reported that more than 8,000 people in Malaysia had been admitted to hospital with health problems related to the pollution.

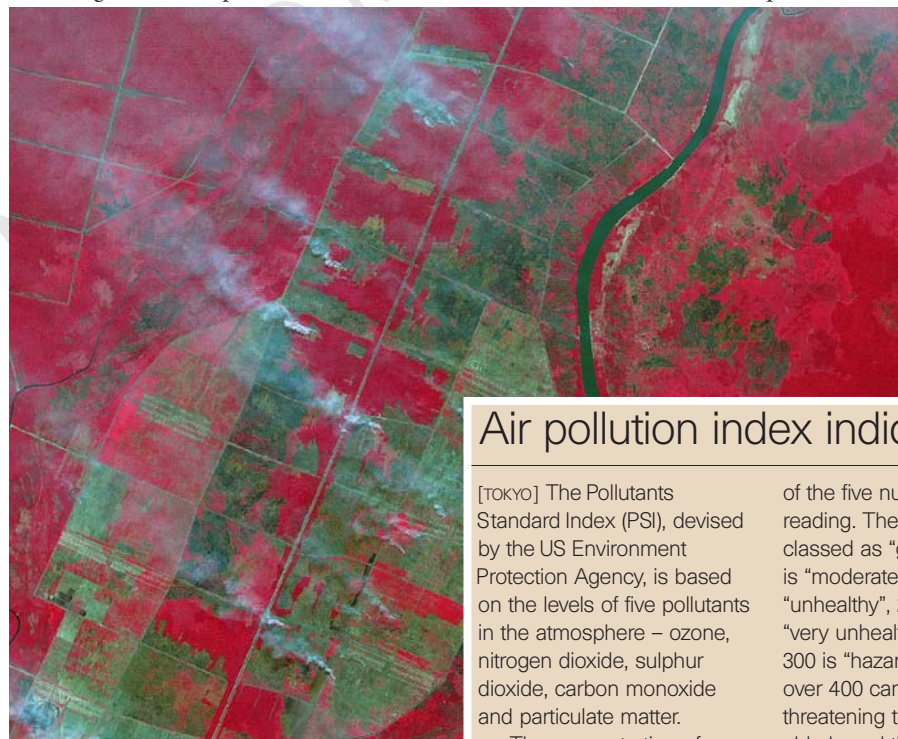
The Environment Ministry in Singapore has announced that if the Pollutants Standard Index, a measure of the level of various pollutants in the atmosphere devised by the US Environment Protection Agency (see box), reaches the "hazardous" level of 300, schools and sports complexes will be closed down. If it reaches the "life-threatening" level of 400, civil defence sirens will sound and emergency radio broadcasts will be made.

Singapore has not so far been as badly affected as Malaysia and Indonesia because winds are blowing most of the smoke into Malaysia. Nevertheless, the index reached an all-time high of 226 in Singapore for a 3-hour period last Thursday, a level classed as "very unhealthy" under the US measure.

Malaysia uses a slightly different scale for measuring pollution called the Air Pollution Index. But it is based on the same principles and yields similar though slightly higher numbers to the Pollutants Standard Index (see panel below).

In Sarawak, the Air Pollution Index hit 658 in Kuching when the government declared a state of emergency. This may be the highest level of pollution ever recorded, as it exceeds the highest level of 500 on the US scale.

David Swinbanks



Burning issue: damage to vegetation in the Indonesian forests has been assessed from satellite pictures analysed by CRISP at the National University of Singapore. (Distributed by SPOT IMAGE/SPOT ASIA)

Air pollution index indicates danger levels

[TOKYO] The Pollutants Standard Index (PSI), devised by the US Environment Protection Agency, is based on the levels of five pollutants in the atmosphere – ozone, nitrogen dioxide, sulphur dioxide, carbon monoxide and particulate matter.

The concentration of each is measured and converted to a number on a scale of 0 to 500. The highest

of the five numbers sets the reading. The range 0 to 50 is classed as "good", 51 to 100 is "moderate", 101 to 200 is "unhealthy", 201 to 300 is "very unhealthy", more than 300 is "hazardous" and levels over 400 can be life-threatening to the ill and elderly and those who have respiratory problems such as asthma.

Malaysia uses a slight

modification of this index called the Air Pollution Index. The scale is the same as the PSI up to 50, but then diverges to slightly higher numbers. For example, a PSI of 200 converts to about 257 on the Malaysian scale and 500 to about 577. Sarawak, where a state of emergency has been declared, recorded a maximum of 658 last Friday.

D.S.

CNES

Jewish leaders seek genetic guidelines...

[SAN FRANCISCO] US Jewish leaders are asking officials of the National Human Genome Research Institute (NHGRI) to discuss the possible development of working guidelines for genetic research on Ashkenazi Jews. They are alarmed by the sentiments reflected in recent newspaper headlines, such as one in New York *Newsday* which referred to bias against "mutant gene carriers".

The current level of research attention by geneticists "feels uncomfortable to the Jewish community", says Amy Rutkin, director of American affairs for Hadassah, the United States' largest Jewish membership organization. "But we understand how important this life-saving research is."

Geneticists at Johns Hopkins University in Baltimore, Maryland, recently reported that they had found a genetic mutation in one of about 17 Ashkenazi Jews — those of Eastern European descent — that seems to double the risk of colon cancer (see *Nature Genetics* 17, 79–83; 1997). The research group says the mutation may be the most common cancer gene found so far within a particular population.

Their study adds to a growing list of mutations in the Ashkenazi population linked to disease, including Tay Sachs, Gaucher's and the 185delAG mutation associated with breast and ovarian cancer.

Such findings, which have already led to Jewish groups being targeted as a potential market for commercial genetic tests, could create the perception that Jewish people are unusually susceptible to disease, says Rutkin.

As a result, she warns, anyone with a Jewish-sounding last name could face discrimination in insurance and employment as companies struggle to keep down health-care costs. Hadassah has been lobbying for laws to forbid genetic discrimination. "Social stigmatization is a profound issue for our community," Rutkin says.

History reinforces such fears. The eugenics movement in the 1920s viewed Eastern and Southern Europeans as genetically inferior in intelligence and prone to criminality. In the United States, these views spurred laws that limited immigration by Jews, other Eastern Europeans and Italians.

For NHGRI, Leslie Fink confirms that discussions are taking place about a formal dialogue with representatives of the Jewish community. She says the institute recognizes that it has a responsibility to combat stigmatization based on genetic difference, and is discussing how to deal with the problem.

Although every individual carries an equal number of genetic mutations that could lead to disease, more is known about the mutations prevalent in certain populations — including Ashkenazi Jews, Finns and

Mormons — partly because researchers have found in these communities relatively convenient to study.

The main reasons are that they are identifiable as a genetically linked population and they tend to keep good genealogical records. As a result, it is much easier to track the presence of a particular mutation, its frequency and its link to the development of disease than it is among undocumented populations.

Another factor is that Jewish people in the United States have long been involved with genetic studies out of both commitment to public service and an awareness of the potential medical benefit to themselves. Genetic screening for Tay Sachs in this community is often held up as an example of the benefits such testing can bring.

But Harry Ostrer, a geneticist at New York University Medical Center who took part in the colon cancer research, says that some religious leaders, concerned about

stigmatization both outside and within the Jewish community, have recently said that they would prefer genetic research on the Ashkenazi population to be abandoned altogether.

Ostrer warns that the premature introduction of genetic tests based on new findings could help to create false impressions and stimulate the discrimination that community activists fear. Hadassah wants to head off any possible backlash against genetic research among Ashkenazi Jews by educating both sides, Rutkin says. She praises NHGRI officials for helping to place recent findings in context.

But Rutkin says it is important for both researchers and members of the community to understand the full implications of genetic research. Scientists must also be more aware, she says, of the ways in which antisemitic notions of the past can become reflected in the interpretation of genetic research today.

Sally Lehrman

... as Israel offers to test adopted children

[JERUSALEM] Israel's Ministry of Science has offered to fund DNA tests for 1,000 Israelis of Yemenite extraction to resolve a long-standing controversy about the fate of children who may have been adopted without their parents' knowledge or consent.

In the early 1950s, a number of Yemenite parents who had recently immigrated to Israel from Yemen and who were living in poverty in transit camps were told that children of theirs who had been hospitalized had died. Often, the parents were taken to their children's supposed graves. Many suspected, however, that the children had been adopted by wealthier families.

These suspicions are supported by testimony heard by a government commission investigating the affair. Earlier this year, an Israeli-born woman now living in the United States travelled to Israel for a DNA test and claims to have been reunited with her biological parents as a result.

In a proposal to the Knesset's special committee

on research and science and development affairs, Adam Friedman, director of the Molecular Biology and Genetic Engineering Unit at Hadassah University Hospital, said that family relationship could be determined by comparing the alleles of several DNA markers in adopted children with those of their supposed biological parents or siblings.

If a given allele is rare among Yemenite Jews, it can provide a strong indication of genetic relationship between two people who carry it, Friedman explained. The greater the number of such rare alleles that are found to be shared by two people, the greater the probability that they are close relatives, he added.

Yemenite Jewish groups now want the government to conduct such tests on all the affected families and on children who were adopted during the early 1950s. Friedman estimates that such a large-scale test would cost about \$200,000.

Mordecai Bashari, the director-general of Israel's Ministry of Science, who is a

physicist by training and is himself of Yemenite extraction, has described the sum as a "ridiculously small amount" and has offered to find it from his ministry's budget.

But although 435 Yemenite Jewish families have applied to the science committee requesting DNA testing, only 35 of the 830 adopted individuals identified by the investigative commission as connected to the affair have asked to be tested. Dalia Itzik, chair of the committee, points out that there is no legal way to force them to be tested.

The families involved also want the putative graves of their children to be opened and any bones found to be tested.

But Yehuda Hiss, director of the Pathological Institute of the Ministry of Health, says that none of Israel's genetic laboratories has enough experience in extracting and analysing DNA from bones to obtain results that would be acceptable as evidence in court, so they would have to be sent abroad for analysis.

Haim Watzman

Creationist appeal told that Ark claims 'were scientific'

[SYDNEY] A Sydney court was told last week that public lectures by Bible church elder Allen Roberts on his claims that he had found evidence for the existence of Noah's Ark in Turkey were scientific, not religious, in character, and should be judged as such.

David Bennett QC was presenting an appeal before three judges in the Federal Court of Australia on behalf of Ian Plimer, a professor of geology at the University of Melbourne. Plimer had accused Roberts of misleading or deceptive conduct towards consumers, an offence for people in business.

The original trial judge, Ronald Sackville, had found Roberts was wrong in his scientific claims and would have breached fair trade acts if he had acted in trade; his verdict was that this was not proven (see *Nature* 387, 540 & 837; 1997). Sackville also said that publicity surrounding the case presented it as a conflict between science and religion, but resolving such matters had no place in court.

After the court hearing, Plimer said he had used his "last shred" of money to finance the appeal and would go bankrupt if he loses. Judgement is expected by the end of the year.

Bennett, who was not involved in the earlier trial, claimed that Sackville had erred in drawing various conclusions under the Fair Trading Acts of the various Australian states and the federal Trade Practices Act.

Summarizing evidence which he claimed proved that Roberts and supporters in the Noah's Ark Research Foundation (NARF) had acted *in* trade or commerce when selling tickets, tapes and publications, Bennett said the goal had been to raise A\$1.95 million (US\$1.4 million) to investigate the site of the supposed remains of the Ark.

Malcolm Duncan, a barrister acting for Roberts, described Plimer's claim that Roberts was acting in trade or commerce as "fundamentally silly" as it "would make the sale (whether for profit or not) of any religious tract actionable under the Fair Trading Acts".

Challenged by Judge Catherine Bransden, Duncan appeared to concede that NARF was in trade. Bennett claimed this as support for the appeal, but the presiding judge, John Davies, did not see it as a concession.

After the original verdict, Roberts claimed that Sackville's verdict "preserved the free speech of anyone who has something important to say publicly". He did not attend the appeal hearing.

Roberts had originally sued Plimer for defamation after expelling him and other geologists from his 1992 lectures for trying to ask questions. This case has yet to be heard. Plimer said he pursued his present action to forestall the libel suit.

Peter Pockley

Scientists criticize review of Indian research institute

[NEW DELHI] A proposal by an international review committee that the Tata Institute of Fundamental Research (TIFR) in Bombay (now Mumbai) should streamline its administration and restructure its research so as to become globally competitive has been sharply criticized by some faculty members, who say it was a rushed job.

The institute was set up by the late Homi Bhabha in 1945. It became a launch pad for India's space and atomic energy programmes, as well as catalysing the growth of its electronic and computer industry. Bhabha also created groups in radioastronomy and computer science which had creditable successes — India's first electronic computer was built at TIFR in the 1960s.

But the country's leading science institute has recently been in the news for less estimable reasons — such as plagiarism by a senior faculty member — and its international reputation has slipped. This prompted its council of management to seek advice from experts abroad.

The review committee was headed by Lord Porter, Nobel prizewinner and past president of the Royal Society, now at Imperial College, London. Other members were Sir Arnold Wolfendale, former Astronomer Royal and president of the Royal Astronomical Society, now at the University of Durham, UK; David Mumford, president of the International Mathematical Union, now at Brown University, Rhode Island; Sydney Brenner of the Molecular Science Research Institute, La Jolla, California; and B. V. Sreekantan, previous director of TIFR.

M. G. K. Menon, a former TIFR director and now a member of TIFR's council of management, admits that the committee's five-day review period in January might have been too short, but says that it had nevertheless done a fine job.

But the report, which was released this month, while being generally complimentary about the state of research, has evoked mixed reactions. "In general, we welcome the report, even though some faculty members are unhappy," says Sudanshu Jha, who has headed the institute since July 1997, and who hopes that the recommendations will help him in his plans to "rebuild" TIFR.

Jha declines to comment, however, on the committee's controversial suggestion that TIFR should give up nuclear energy research and transfer its only accelerator facility — the pelletron — to the Bhabha Atomic Research Centre. The committee recommends that research in solid-state physics, radioastronomy, computer science and molecular biology should be augmented. The institute's high-



Tata institute: abandon nuclear energy research and transfer its accelerator, the review proposes.

energy research programme, though insignificant in world terms, "should continue", but the astronomy programme should be reduced.

The work of the pure mathematics group came in for praise. But a suggestion that the theoretical computer group be merged with pure and applied mathematics has provoked strong opposition. The committee withdrew a similar idea to merge the molecular biology group in Mumbai with TIFR's National Centre for Biological Sciences in Bangalore, after some faculty members threatened to resign rather than move from Mumbai.

In response to the institute's constant complaint of inadequate funding, the committee suggests that it enlist the help of foundations, rather than seek government support, and also increase the commercial exploitation of research results. Although fundamental research is TIFR's main role, it said, the "application of research results to national needs should not be ignored".

The committee also asked the management council to set up a panel to examine staff promotion policy and the possibility of a recruitment drive worldwide. It described the shortage of people joining the institute as "surprising and worrying". Menon claims that TIFR has already set up a group to revamp the administration and to look into promotion and recruitment policies.

"The report is not a court order," says Jha, who has distributed copies to all staff, asking for their comments. "Whatever suggestion is practical, we will implement."

But Nataraja Sarma, a nuclear physicist associated with TIFR for 35 years, says that two earlier internal reviews produced no improvement, and that entrenched interests in the institute could mean that although "the present review is reasonable, no serious changes will ensue".

K.S. Jayaraman

Australia rallies Pacific states against binding greenhouse targets

[LONDON] Australia is to lead a bloc of south Pacific countries which will oppose legally binding targets for greenhouse gas emissions at this year's United Nations annual climate convention in December.

A meeting of the 16-member South Pacific Forum last week endorsed Australia's call for individual nations to set their own targets. John Howard, prime minister, said concern over rising sea levels was exaggerated. The meeting had been called to decide a regional strategy for the UN meeting in Kyoto, Japan.

Australia argues that a blanket cut on emissions will unfairly penalize the country, 90 per cent of whose electricity is coal-generated. Environmentalist groups say Australia is one of the world's highest polluters per head of population.

Environmentalists' ally within DOE quits

[WASHINGTON] Tara O'Toole, assistant secretary of energy for environment, safety and health and a staunch ally of environmental groups in the Clinton administration, has resigned. "It is time for

me to step back, rest and reflect," she says.

O'Toole, a medical doctor who worked for the congressional Office of Technology Assessment before joining the administration in 1993, was the driving force behind an effort to expose unethical human subjects research conducted by the US government. More recently, she led a health and safety crackdown at the Brookhaven National Laboratory in New York state, where the contractor was fired after the discovery of a leak of radioactive tritium.

Federico Peña, the energy secretary, issued a statement expressing his "deep regret" at O'Toole's sudden departure. "She embodies the best in public service," he said. "She has made a difference."

Telescopes survey infrared sources

[WASHINGTON] A comprehensive telescopic survey of the entire infrared sky began last week, with the goal of updating a 30-year-old catalogue of near-infrared astronomical sources. The Two-Micron All-Sky Survey (2MASS), led by the University of Massachusetts and funded by the National Aeronautics and Space Administration and the National Science Foundation, will last for three and a half years.

A pair of dedicated 1.3-metre telescopes observing from Mount Hopkins, Arizona

and Cerro Tololo, Chile, will photograph the sky and catalogue an estimated one million galaxies and 300 million stars as well as quasars, asteroids, and brown dwarfs. The telescopes will use detectors developed for the Hubble Space Telescope's Near Infrared Camera and Multi-Object Spectrometer (NICMOS). 2MASS is expected to be 25,000 times more sensitive than a similar infrared sky survey completed in 1969.

'Eugenic' sterilizations admitted by Japan

[TOKYO] The Japanese government has acknowledged that 16,000 disabled women were sterilized without consent between 1949 and 1995 under the Eugenic Protection Act of 1947. Seventeen groups representing women and disabled victims compelled the government to carry out an investigation, following similar revelations in Sweden. They also requested an official apology and compensation, but the government has so far shown no intention of responding.

The law was passed to authorize abortions and sterilizations with eugenic objectives. It was abolished recently on the grounds that it was a violation of human rights. The announcement was precipitated by recent controversy in Sweden, where 60,000 Swedes were involuntarily sterilized over a period of 40 years, until 1976.

Clinton threatens to veto NIH budget bill

[WASHINGTON] US president Bill Clinton has threatened to veto a government spending bill in the House of Representatives that raises the budget for the National Institutes of Health (NIH) by six per cent to \$13.5 billion. Clinton said last week that he would veto the \$279 billion bill funding the Departments of Labor, Health and Human Services and Education in 1998 unless Republican-inspired education-related amendments and spending cuts opposed by the president are withdrawn.

The fiscal year ends on 30 September. If no 1998 bill has been enacted by then, a temporary spending measure is expected to be introduced to keep the NIH budget at its current levels. A companion Senate bill increases NIH funding by 7.5 per cent.

'Rules treat students as economic refugees'

[MUNICH] The German Academic Exchange Service (DAAD) and the German Rectors' Conference (HRK) have criticized Germany's interior minister for tightening restrictions on overseas students at a time when science officials have become concerned at the diminishing number of students from overseas who want to study in Germany.

Under proposals suggested by Manfred Kanther, the interior minister, overseas students would need to prove in advance that they have sufficient funds to complete their studies before being admitted to study in Germany. But Theodor Berchem and Klaus Landfried, the presidents of the DAAD and HRK, describe the proposals as "unnecessary barriers", which will in turn discourage students from wanting to study in Germany. They say that the recruitment situation is unlikely to improve if students are treated as "economic refugees".

NASA attacked over Mir mission risks

[WASHINGTON] Congressional critics have called on the National Aeronautics and Space Administration (NASA) to stop sending its astronauts to the accident-prone Russian Mir space station. In an emotional hearing last week, James Sensenbrenner (Republican, Wisconsin), chairman of the House Science Committee, said: "The risks on board Mir have increased while the scientific benefit has decreased." Sensenbrenner asked Daniel Goldin, NASA's administrator, who was not present, to reevaluate his decision to leave the NASA astronaut David Wolf on the station for a four-month stay.

Sensenbrenner and other committee

members even invoked the memory of the 1986 disaster of the space shuttle Challenger, asking "Does someone have to get killed?" NASA, however, insists that Mir is safe and has proceeded with plans to launch Wolf and his crewmates into orbit this week. The agency is under no obligation to follow Sensenbrenner's wishes. But the hearing is likely to worsen the already strained relations that exist between Goldin and the committee's chairman.

Israeli ministry 'running out of research funds'

[JERUSALEM] Israel's Ministry of Commerce and Industry has warned that it will be unable to provide any new research and development grants this year if the treasury fails to provide additional funding. According to a ministry spokesperson, a shortage of grants could cause Israeli high-tech industry to transfer its research overseas.

The ministry is currently negotiating with the Ministry of Finance for additional funding for this year, and is opposing cuts planned for next year. Ministry officials say that their budget for research and development is \$450 million short of what is needed to fund any further grants this year, and there will be a greater shortfall next year if the proposed budget cut is maintained.

Fat (and thin) rats distort results

Sir — It was gratifying to see M. Festing¹ address one of the most vexing problems in toxicology, that of rodents so obese that the validity of carcinogenesis screening is brought into question^{2,3}.

Festing recommended the use of isogenic rather than outbred strains of rat. But he did not mention some work that had already been done using isogenic strains. The problem in carcinogenicity screening tests of increasing tumour rates (in the anterior pituitary gland, for example) and decreasing survival, accompanied by an increase in body weight, was actually first reported in an isogenic strain, the F-344 rat⁴.

We have related body weights to survival and neoplasia (such as leukaemia) in the same F-344 rat⁵. We⁶ and others⁷ have also correlated body weights to liver (and other) tumour incidences as well as increases in body weights to increases in tumour incidences over the same time period in another isogenic strain used in carcinogenicity tests, the B6C3F1 mouse. These studies have identified a major source of the variability in the tumour incidences of control animals used in carcinogenicity tests as the differences in body weights in the various studies.

Rather than incurring the problems that ensue from changing strains, or trying to

breed small animals, one approach is to use dietary control⁶, which maintains body weights at predetermined practical ranges through control of dietary intake.

This approach not only prevents the 'fat rat', but also minimizes the variability in tumour endpoints and survival associated with the usual wide range in body weight seen in toxicology studies. Control of this variability is important because, besides being too large, rodents can also be too small in carcinogenicity bioassays, because of either an inadvertent restriction of feed or exposure to an agent, and thus be fairly insensitive to the action of toxicants.

We strongly support the evaluation of new models in biological studies. However, new models require painstaking evaluation and comparison with those now accepted by the toxicological community so that their results can be put into a proper context. We have a wealth of results in the strains used in toxicity testing that allow the kinds of comparison necessary for the development of new procedures that can provide improved means to assess human risk.

The complicating effects of uncontrolled body weight in toxicity tests is a consequence of the practice, common in rodents although unusual for most test species, of *ad libitum* feeding.

These complications do not appear

restricted to either isogenic or outbred strains, or only to certain rodent species. Indeed, they are not confined to carcinogenicity tests *per se*. The doses required to produce short-term toxicity vary two- to fourfold, based on the dietary intake and resulting body weight of the animal⁸.

So dietary control could be a better solution than the genetic approach to resolve the problems currently observed in toxicity tests resulting from 'fat rats' and the other sequelae of uncontrolled food consumption and growth.

Angelo Turturro

Julian Leakey

William T. Allaben

Ronald W. Hart

National Center for Toxicological Research,
3900 NCTR Road,

Jefferson, Arizona 72079, USA

e-mail: aturturro@nctr.fda.gov

1. Festing, M. F. W. *Nature* **388**, 321–322 (1997).
2. Keenan, K. et al. *The Carcinogenicity Debate* (eds McAuslane, J., Lumley, C. & Walker, S.) 77–102 (Quay, London, 1992).
3. Hart, R. W. et al. *Science* **270**, 1419–1421 (1995).
4. Rao, G. et al. *Toxicol. Path.* **18**, 61–70 (1990).
5. Turturro, A. & Hart, R. *Biological Effects of Low-Level Exposures: Dose-response Relationships* (ed. Calabrese, E.) 143–152 (Lewis, Michigan, 1994).
6. Turturro, A. et al. *Toxicol. Path.* **24**, 769–775 (1996).
7. Seilkop, S. *Fund. Appl. Toxicol.* **24**, 247–259 (1995).
8. Keenan, K. et al. *Toxicol. Path.* **22**, 300–315 (1994).

Whodunnit?

Sir — A recent leading article detailed some of the difficulties involved in arbitrating authorship of scientific papers and concluded that attempts to define rules for authorship are "doomed to fail" (*Nature* **387**, 831; 1997). Although this is probably true, some of the thornier issues surrounding authorship could be mitigated if journals simply required the contribution of each author to be briefly stated. The rules the authors had applied in determining authorship would then at least be explicit.

For example, a paper (similar to that described in the leading article) on the discovery of a previously unknown hominid using a novel fossil detector might include the following statement: CM: fossil discovery, morphometry, principal author; PW: stratigraphy, assessed paper; MC: isotopic dating; LG: inventor of CRD (carbon replacement detector); PC: funding, intellectual contributions, co-authored paper.

Such a statement could be placed in the acknowledgements and would serve to allocate both credit and responsibility for the work being published. Such information, if

generally available, would be widely useful in making hiring and tenure decisions, evaluating grant applications, judging published work submitted for doctoral theses, and, when necessary, determining responsibility for fraud.

Despite its utility, some investigators are likely to view this proposal as unnecessary, inconvenient or even demeaning. Journal editors will want to tread lightly so as not to alienate their contributors. Nevertheless, by taking the longer view (as you encourage authors to do), journals can serve the enterprise of science by helping authors to allocate credit fairly for their work and to accept responsibility for it.

Nature's pre-eminent position among journals of science gives it a unique opportunity in this respect. Few prospective contributors, I suspect, would forgo the opportunity of publishing in *Nature* simply because they were required to state their respective contributions.

Benjamin White

Department of Pharmacology,

Yale University,

School of Medicine,

New Haven, Connecticut 06520, USA

e-mail: bhwhite@biomed.med.yale.edu

Nepotism and sexism in peer-review

Sir — I and my colleagues read with interest the Commentary by Christine Wennerås and Agnes Wold about nepotism and sexism in peer-review in the Swedish Medical Research Council (*Nature* **387**, 341–343; 1997).

We keep under constant review the fairness of our peer-review procedures and seek to ensure that our boards, committees and panels are constituted to reflect the clinical, non-clinical, gender, age, ethnic origin and geographical spread of the biomedical community in the United Kingdom.

But although our own data (which are available on request) do not suggest *prima facie* evidence of bias, the Swedish study has prompted us to carry out a review of the outcomes of our competitions so that we can be sure we have done all we can to eliminate bias.

Gillian Breen

Medical Research Council,

20 Park Crescent,

London W1N 4AL, UK

It's not easy to make the Moon

Jack J. Lissauer

Theory has it that the Moon grew within a disk of material splashed out of the Earth by a body the size of Mars. According to new calculations, however, the impacting body was at least twice that size.

Earth's Moon is one of the most peculiar bodies in the Solar System. It is slightly more than 1 per cent as massive as the Earth, its primary, whereas (excluding the tiny Pluto–Charon 'double planet') all other moons are less than one-fortieth of 1 per cent as massive as their primaries. Moreover, the Moon's composition is quite unusual — it is an extremely dry place, devoid of water and other volatile substances; yet it is anomalously low in density for a body in the inner Solar System, because it is depleted in iron relative to a solar-composition mix of refractory (high-condensation-temperature) elements.

Presumably, the satellites of the giant planets formed within disks of gas and dust orbiting their primary (as did the planets themselves, the primary in this instance, of

course, being the Sun). But it is unclear how a solid planet such as Earth could acquire such a disk, so other models of lunar formation have been developed. Each, however, has had its drawbacks.

Was the Moon born from the Earth — that is, thrown off from a rapidly rotating young planet (the fission hypothesis)¹? If most of Earth's iron had already settled into the core, this model could explain the Moon's severe anaemia, but the dynamical aspects of the theory seem intractable². Did the Moon grow from a disk of planetesimals captured into Earth orbit (the coaccretion hypothesis)³? It appears unlikely that a solid planet would acquire a massive disk in such a manner, and this model does not account for the Moon's low iron content. Finally, was the Moon captured into orbit about the Earth,

either intact or in fragments torn apart by Earth's tidal pull? This is unlikely according to dynamical calculations², and such a model again doesn't explain the Moon's low abundance of iron, which is one of the most common elements that condenses at relatively high temperatures.

All in all, developing a theory of lunar origins that could make sense of data obtained from the Apollo lunar landing programme proved very difficult. So much so, in fact, that when I took a class on our planetary system from Irwin Shapiro two decades ago, he joked that the best explanation was observational error — the Moon does not exist.

Around that time a new theory was proposed, one which relied on a giant off-centre impact by an exceptionally large planetesimal to eject material from Earth's mantle into terrestrial orbit, from which it could have grown into the Moon^{4,5}. Such an impact would generate a massive cloud of silicate vapour, which could provide the forces needed to place much of the ejecta into stable orbits about the Earth. This model appeared to satisfy the basic dynamical and chemical constraints of the problem, and it won widespread acceptance at the 'Origin of the Moon' conference in 1984 (ref. 6).

Subsequent modelling of the impact supported the basic scheme of events (although it now seems that gravitational torques were more important than gas pressure in preventing the impact ejecta from falling back to Earth in less than one orbit, and that the mantle of the impactor contributed a substantial fraction of the mass ending up in terrestrial orbit). Modelling also revealed that the impactor would need to be at least comparable in mass to the planet Mars^{7,8}, that is, about one-tenth as massive as the Earth.

Only recently, however, has the process of lunar accretion in such an impact-generated disk been simulated in any detail, which is where a paper by Ida, Canup and Stewart (page 353 of this issue⁹) comes in. This is by far the most detailed simulation of lunar growth in an impact-generated disk published to date. Ida *et al.* use direct *N*-body simulations to show that a single dominant moon is usually able to grow from such a disk. However, most of the disk material orbiting near or interior to Roche's tidal limit of accretion at the beginning of the simulation is lost to the planet. (Tidal forces increase rapidly as one approaches a planet; these tides can be strong enough to rip a body apart, and Roche's tidal limit is the location inside which a satellite held together only by its self-gravity will be torn apart by the tidal force of its primary.) The implication here is that lunar growth in an impact-produced disk is not very efficient. So, to form our Moon, more material must be placed in orbit at a greater distance from Earth than was previously believed.

Simulations by Cameron⁸, published



Figure 1 Earth and its Moon, as imaged from the Galileo spacecraft. There are probably very many terrestrial planets in our Galaxy, yet the implication of the latest simulations^{8,9} is that fewer of them than previously expected have Moon-sized satellites.

earlier this year, dealt with protolunar disk formation. If both his results and those of Ida *et al.* prove correct, they suggest that the impactor needed to form a disk large and massive enough to create the Moon would have had to have been more than twice as massive as Mars, and that it must have provided the Earth–Moon system with a few times as much angular momentum as the system currently possesses. This presents two difficulties.

First, how was this ‘excess’ angular momentum lost (solar tides have removed a small amount, but far less than required)? If the Earth was rapidly rotating before the impact, this problem might be less severe, but such simulations have not yet been performed. Second, simulations of the late stages of terrestrial-planet growth¹⁰ imply that an Earth-sized planet is fairly likely to accrete a body roughly as massive as Mars during its growth, so a substantially off-centre impact of a body this size could occur during the growth of tens of per cent of Earth-like bodies, and Moon-like companions to Earth-like planets might be

fairly common. But if a much more massive impactor is required, then the Earth–Moon system may be something of an anomaly. The only extrasolar terrestrial planets thus far discovered orbit pulsars, but according to theory there are billions of such planets in our Galaxy. If the new simulations are correct, most of these planets do not have Moon-sized satellites unless they form by some mechanism other than a giant impact. □

Jack J. Lissauer is in the Space Science Division, NASA Ames Research Center, Moffett Field, California 94035, USA.

1. Darwin, G. H. *Phil. Trans. R. Soc. II* **170**, 447–538 (1879).
2. Boss, A. P. & Peale, S. J. in *Origin of the Moon* (eds Hartmann, W. K., Phillips, R. J. & Taylor, G. J.) 59–101 (Lunar & Planetary Institute, Houston, TX, 1986).
3. Ruskol, E. L. *The Moon* **6**, 190–201 (1973).
4. Hartmann, W. K. & Davis, D. R. *Icarus* **24**, 504–515 (1975).
5. Cameron, A. G. W. & Ward, W. R. *Lunar Sci.* **VII**, 120–122 (1976).
6. Hartmann, W. K., Phillips, R. J. & Taylor, G. J. (eds) *Origin of the Moon* (Lunar & Planetary Institute, Houston, TX, 1986).
7. Benz, W., Slattery, W. L. & Cameron, A. G. W. *Icarus* **71**, 30–45 (1987).
8. Cameron, A. G. W. *Icarus* **126**, 126–137 (1997).
9. Ida, S., Canup, R. M. & Stewart, G. R. *Nature* **389**, 353–357 (1997).
10. Wetherill, G. W. *Science* **228**, 877–879 (1985).

Allosteric proteins

Cuddling up to channel activation

Christopher Miller

Proteins are intelligent beings. They have evolved to operate in the metabolic maelstrom of a turbulent cellular environment. Transcription factors must know when to switch genes on or off, and the cellular levels of specific ‘signalling’ molecules — lactose, retinoic acid, tryptophan or copper, to mention a few cases — give them this information. Likewise, enzymes at key biochemical control points have to speed up or slow down according to the ever-changing demands, coded in concentrations of cytoplasmic metabolites, of, well, life. Haemoglobin, the granddaddy of all such ‘allosteric’ proteins, knows when you are sleeping or sprinting, realizes whether you live in Cape Cod or Kathmandu, and ascertains moment by moment whether it is coursing through your lungs or visiting vigorously respiring tissues; it makes these judgements and accordingly adjusts its conformation, and thus the blood’s oxygen-carrying behaviour, by sensing cellular solutes such as CO₂, H⁺, Cl[−], NO and bisphosphoglycerate.

On page 389 of this issue, Ruiz and Karpen¹ give us an unprecedented look at the workings of an allosteric membrane protein, the cyclic nucleotide-gated (CNG) ion channel. Their experiments provide a close-up view of channel activation by cyclic GMP and also address the biologically pervasive issue of allostery: how the binding of small molecules leads to conformational changes in proteins.

The CNG cation-conducting channels are central elements in the conversion of sensory input (light, odour, taste) into electrical signals in neurons^{2,3}. The channels are activated by binding intracellular cGMP or cAMP, themselves controlled by an unceasing tug-of-war between nucleotide cyclases and phosphodiesterases which respectively synthesize and break down cyclic nucleotides. The channel studied electrophysiologically by Ruiz and Karpen is the cGMP-activated channel of mammalian rod photoreceptors expressed in frog oocytes. As cGMP increases, the channel turns on in a cooperative fashion^{2–4}. A measure of the steepness of this S-shaped activation curve, and hence of the quality of the channel as a molecular switch, is the Hill coefficient, which on thermodynamic grounds is also a lower limit on the number of cGMP binding sites per channel. Typically, CNG channels give Hill coefficients as high as three, a value harmonious with their tetrameric architecture^{4,5}.

Because the channel carries four identical cGMP-binding sites, it is natural to ask how many sites must be occupied before the channel opens. With four subunits to keep track of, things get complicated quickly, even assuming minimally that each subunit can exist in either of two shapes — closed-like (C) and activatable (A). One well-recognized possibility is that cGMP binding to a given subunit promotes a C→A conforma-

tional change in that subunit alone, but that the channel opens only when all four subunits simultaneously adopt the A form.

Alternatively, the Monod–Wyman–Changeux (MWC) model⁶, so popular because of its algebraic tractability (and despite its macromolecular implausibility), has it that any conformational change of the channel must retain symmetry; only C₄ or A₄ may exist, with the A₄ form presenting four identical cGMP-binding sites of higher affinity than the C₄ form. More complicated ways of linking ligand binding to the confor-

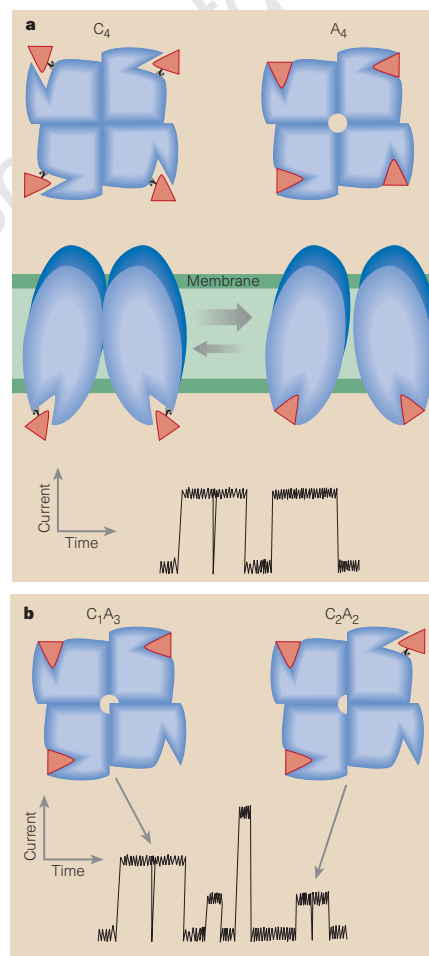


Figure 1 The innovative techniques used by Ruiz and Karpen¹ enable them to analyse not only the behaviour of single cyclic nucleotide-gated (CNG) ion channels, but also the effects of ligands locked onto each of the channel's four subunits. a, Top and lateral representations of a CNG channel with the cyclic GMP ligand molecules covalently confined to the four available activation sites. Two symmetrical conformational states of the channel are shown, fully closed (C₄) and fully open (A₄), each with its own icon. b, A possible scheme to explain the partial openings, or ‘substate’ conductances, observed in single-channel records from triply liganded channels. Simulated channel recordings are shown for the closed and fully open states, as well as for substates putatively corresponding to channels of mixed conformation (indicated by arrows).

mational change of the subunits can be imagined, and the difficulty of experimentally discriminating among them usually provokes a mind-numbing denial: let's just use the MWC model heuristically and hope for the best!

This is a reasonable attitude when it is experimentally impossible to unravel the combinatorial morass of molecular forms present: liganded and unliganded subunits in either conformation, combined in a tetrameric complex. Indeed, the elusive 'intermediate forms' have tormented haemoglobinologists for over 70 years⁷⁻⁹.

Ruiz and Karpen¹ now cut through this tangle by combining two powerful techniques in an experimental *tour de force* that allows a huge simplification in analysis. First, they apply inside-out patch recording to observe single CNG channels and thus to avoid the ensemble-averaging intrinsic to any test-tube measurement. Second — and this is the jim-dandy innovation — they isolate pure forms of the channel in their recording system: channels with exactly one, two, three or four cGMP molecules locked onto their binding sites. This they accomplish by photochemically tethering a cGMP analogue to the single channel¹⁰; by varying light exposure, they prepare individual channel proteins with known numbers of sites covalently occupied. So channels with fixed ligand occupancies can be observed directly without any cyclic nucleotide in the bathing solution (Fig. 1a).

The results are striking. Quadruply occupied channels stay open nearly all the time in a high-conductance state. Unliganded channels, observed previously^{11,12}, and those with one ligand, open with a minuscule but readily measured probability ($\sim 10^{-5}$). Doubly and triply occupied channels open $\sim 1\%$ and 30% of the time, respectively, but the openings frequently show 'substate' conductances lower than the large conductance seen with four ligands. (Taylor and Baylor¹³ had observed substates at low cGMP levels and proposed that they represent partially liganded channels, as is now verified directly.)

Do the substates represent non-symmetric tetramers of mixed conformation, such as C_1A_3 or C_2A_2 , yielding 'partial openings' of the pore (Fig. 1b)? The authors won't stick their necks out in print to identify them as such 'Koshland intermediates'¹⁴, but I'd wager that images like this populate their dreams and provoke future experimental plans.

The experiments offer exciting possibilities for close analysis of allosterically controlled channel gating, and they greatly constrain ideas about CNG channel activation by showing that concerted, symmetric gating models such as MWC are inadequate to represent linkage between cGMP binding and channel opening. One gratifying simplification of the channel's activation pathway is clear, though: only the three- and four-lig-

anded channels contribute significantly to cGMP activation of total conductance. If this were not the case, the channel would be a low-quality, sluggish switch which would not display a 'threshold' cGMP concentration for activation, and our ability to discriminate gradations of illumination, colour or smells would be compromised. The fact that none of these details could have been uncovered without the microscopic experiments described here should make us all wonder what is really happening when ligand binding drives conformational changes in the conventionally studied proteins of biochemistry textbooks. □

Christopher Miller is in the Howard Hughes Medical Institute, Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02254, USA.

Artificial enzymes

The importance of being selective

John T. Groves

Hydroxylating enzymes introduce one or more O-H groups into organic compounds and are involved in many biological processes. But how do they work? And how might these systems be redesigned to practical ends? Bioorganic chemists address these questions by using synthetic molecular systems.

A striking example of such research appears in the *Journal of the American Chemical Society*¹, where Breslow *et al.* describe the catalytic hydroxylation of a steroid using an artificial cytochrome P-450 enzyme. What is remarkable about this report is both the positional selectivity the authors have achieved, and that the reaction exhibits true catalytic turnover of steroid molecules.

Mechanistic and structural studies, and the simultaneous development of metalloporphyrin model systems, has made

1. Ruiz, M. & Karpen, J. W. *Nature* **389**, 389–392 (1997).
2. Zimmerman, A. *Curr. Opin. Neurobiol.* **5**, 296–303 (1995).
3. Zagotta, W. N. & Siegelbaum, S. A. *Annu. Rev. Neurosci.* **19**, 235–263 (1996).
4. Zimmerman, A. & Baylor, D. A. *Nature* **321**, 70–72 (1986).
5. Liu, D. T., Tibbs, G. R. & Siegelbaum, S. A. *Neuron* **16**, 983–990 (1996).
6. Monod, J., Wyman, J. & Changeux, J.-P. *J. Mol. Biol.* **12**, 88–118 (1965).
7. Adair, G. S. *J. Biol. Chem.* **63**, 529–545 (1925).
8. Ackers, G. K. *Trends Biochem. Sci.* **18**, 385–390 (1993).
9. Edelstein, S. J. *J. Mol. Biol.* **257**, 737–744 (1996).
10. Brown, R. L., Gerber, W. V. & Karpen, J. W. *Proc. Natl Acad. Sci. USA* **90**, 5369–5373 (1993).
11. Picones, A. & Korenbrot, J. I. *J. Physiol. (Lond.)* **485**, 699–714 (1995).
12. Tibbs, G. R., Goulding, E. H. & Siegelbaum, S. A. *Nature* **386**, 612–615 (1997).
13. Taylor, W. R. & Baylor, D. A. *J. Physiol. (Lond.)* **483**, 567–582 (1995).
14. Koshland, D. E., Némethy, G. & Filmer, D. *Biochemistry* **5**, 365–385 (1966).

cytochrome P-450 the Rosetta stone among the oxidizing iron enzymes (which also include cytochrome oxidase and methane monooxygenase, a non-haem di-iron protein). Of the reactions mediated by P-450 (ref. 2), the selective insertion of oxygen into the C-H bonds of unreactive molecules such as steroids and prostaglandins is of especial interest. There is no general way to introduce oxygen into C-H bonds in organic molecules. Yet such compounds are often key intermediates in the synthesis of drugs and hormones.

The reaction cycle of P-450 is known to be initiated by substrate binding. A cluster of water molecules is displaced from the enzyme's active site and the electronic configuration of the iron changes in response. So it is only the enzyme-substrate complex that proceeds to bind and activate molecular oxy-

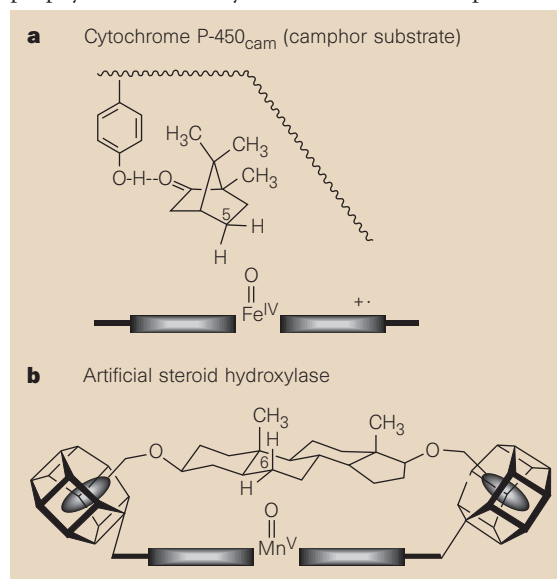


Figure 1 Hydroxylation reactions. a, Cytochrome P-450_{cam} enzyme with the substrate camphor at the active site; b, the system designed by Breslow *et al.*¹. In b, a synthetic manganese porphyrin replaces the haem centre of the natural enzyme, and the substrate steroid is modified with aromatic groups at either end. Most notably, in b hydroxylation occurs with high selectivity at carbon 6 of the steroid.

gen. The central mechanism of all these systems is that a reactive oxo-iron intermediate ($\text{Fe}=\text{O}$) is produced at the active site. Oxygen transfer from this intermediate produces the product alcohol in a process known as oxygen rebound. In the case of model systems, the reactive oxo-iron and oxo-manganese porphyrins have been directly observed. Further, the rates of oxygen transfer from such species, which have only recently been measured, are extraordinarily fast³.

Clues to the origin of site selectivity can be discerned from the X-ray crystal structure of cytochrome P-450_{cam} which has been determined with the substrate camphor at the active site (Fig. 1a)⁴. Carbon 5 of the camphor molecule is found near the haem iron centre of the protein and the carbonyl oxygen of camphor is hydrogen-bonded to a tyrosine hydroxyl supplied by the protein in the active site. From this, one might guess that the tyrosine hydrogen bond would be essential to the very high selectivity of this enzyme: remarkably, however, replacing the tyrosine with phenylalanine by selective mutagenesis only marginally decreased the selectivity⁵.

Inspection of the camphor molecule suggests that the hydrogens at carbon 5 should be intrinsically more reactive due to a combination of steric and electronic effects. Further, the stepwise mechanism of C–H bond cleavage followed by C–O bond formation allows the initial attack to involve either of the two hydrogens at carbon 5. Blocking carbon 5 in camphor by replacing the hydrogens with fluorine led to the hydroxylation of other positions in the same region of the molecule⁶. So the substrate camphor must be able to move within the active site during the lifetime of the reactive iron intermediate to expose other sites to hydroxylation. In this light, the high selectivity of the wild-type enzyme can be seen to result from a combination of intrinsic selectivity, mechanistic determinants and subtle dissuasions by restricting access to some reaction pathways in much the manner of an intramolecular reaction. Indeed, the first evidence for reactive oxo-iron intermediates in model iron porphyrin systems derived from the observation of the hydroxylation of an attached substrate⁷.

Breslow and colleagues' strategy¹ to induce selectivity in their model system was to attach four cyclodextrin appendages to a synthetic manganese porphyrin used in place of the haem centre of the natural enzyme. These doughnut-shaped heptamyllose sugars have a hydrophobic central cavity which is known to bind aromatic molecules. The substrate steroid was modified with such an aromatic group at either end of the molecule. The host–guest complex obtained from these designed partners displays a limited region of the substrate steroid in the vicinity of the catalytic

manganese centre (Fig. 1b).

Most significantly, Breslow *et al.* found that hydroxylation occurred only at carbon 6 to give the 6 α -hydroxysteroid, even though there are many sites in this molecule with similar intrinsic reactivity. This high selectivity depended critically on the precise arrangement of the aromatic groups of the substrate. Moreover, the manganese porphyrin–cyclodextrin construct was able to release the product steroid and hydroxylate at least four steroid molecules. Catalytic turnover with such high positional selectivity is unusual for this kind of model system.

A true artificial enzyme must be able to turn over many times by releasing the product and binding a new substrate. Indeed, product release is often the slowest step in enzymatic catalysis. Single-site hydroxylation of a steroid has been observed before with a membrane-bound manganese porphyrin⁸; here, complete selectivity for the hydroxylation of cholesterol at carbon 25 was induced by the weak intermolecular forces between the metalloporphyrin, the lipid membrane and the steroidal substrate. However, this model system turned over only once.

The multiple turnovers reported by Breslow *et al.*¹ show that this simple system also has the conformational flexibility required to release the product steroid to the medium so that another substrate can bind. Results

such as these can confirm that our understanding of the enzyme mechanism is correct, but they can also point the way to new catalytic processes based on biochemical analogy. Work with ruthenium porphyrins has demonstrated very high reactivities and turnover numbers approaching 1,000 per minute for hydrocarbon hydroxylation⁹. If the high efficiency of this ruthenium system could be combined with the high-selectivity strategies offered by Breslow *et al.*, synthetic enzymes could offer efficient, catalytic routes to many inaccessible drug intermediates. □

John T. Groves is in the Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA.

e-mail: jtgroves@princeton.edu

1. Breslow, R., Zhang, X. & Huang, Y. *J. Am. Chem. Soc.* **119**, 4535–4536 (1997).
2. Ortiz De Montellano, P. (ed.) *Cytochrome P-450* 2nd edn (Plenum, New York, 1995).
3. Groves, J. T., Lee, J. & Marla, S. S. *J. Am. Chem. Soc.* **119**, 6269–6273 (1997).
4. Poulos, T. L., Fingel, B. C. & Howard A. J. *J. Biol. Chem.* **195**, 687–700 (1987).
5. Meuller, E. J., Loida, P. J. & Sligar, S. G. in *Cytochrome P-450* 2nd edn (ed. Ortiz De Montellano, P.) 83 (Plenum, New York, 1995).
6. Eble, K. S. & Dawson, J. H. *J. Biol. Chem.* **259**, 14389–14393 (1984).
7. Groves, J. T., Nemo, T. E. & Myers, R. S. *J. Am. Chem. Soc.* **101**, 1032 (1979).
8. Groves, J. T. & Han, Y.-Z. in *Cytochrome P-450* 2nd edn (ed. Ortiz De Montellano, P.) 3–48 (Plenum, New York, 1995).
9. Groves, J. T., Bonchio, M., Carofiglio, T. & Shalyaev, K. V. *J. Am. Chem. Soc.* **118**, 8961–8962 (1996).

Conservation biology

The hen harrier and the grouse

Robert M. May

In Britain, the hen harrier, *Circus cyaneus*, is a beautiful but scarce bird of prey, today found mainly on moorland in Scotland. In the early 1800s it was both more abundant and more widespread, but loss of lowland habitat and human persecution in the late nineteenth and early twentieth centuries



Figure 1 A female hen harrier and its young. The birds' 'nest success' is much lower on grouse moors than elsewhere.

effectively eradicated it from mainland Britain. Beginning in the 1930s and 1940s, it spread back into parts of upland Britain from Orkney and the Outer Hebrides, and according to the most recent estimates, there are around 600 breeding females in Britain, including 40 on the Isle of Man^{1,2}.

A definitive study of the population dynamics of the hen harrier has, however, been lacking. That deficiency is now remedied by work by Etheridge, Summers and Green, published in the *Journal of Applied Ecology*³, which highlights some of the conflicting forces at play in conservation biology.

The hen harrier's big problem is that it eats chicks and adults of the red grouse, *Lagopus lagopus scoticus*. The grouse-shooting industry is of some economic significance, directly generating a gross revenue of more than £10 million (\$16 million) a year, along with indirect knock-on income⁴. More importantly from a conservation viewpoint, management of moorland estates for grouse has helped protect large areas of Britain's uplands, and their native vegetation, from overgrazing or disappearing under gloomy

plantations of sitka pine or other exotic tree species. But successful management of the grouse-shooting industry — the 'Glorious Twelfth' of August, which marks the start of the shooting season, and all that — depends on high population densities of grouse. So the hen harrier's depredations are understandably viewed with disfavour by moorland managers. Although hen harriers and their nests have received full legal protection since 1954, there are abundant anecdotal reports of their persecution.

Etheridge and colleagues³ have carried out a thorough study of the breeding productivity, natal dispersal and survival of hen harriers over the eight-year span 1988–95. The study covered moorland managed for grouse shooting, other heather moorland and young conifer forests in the Scottish uplands. The authors found 'nest success' much lower on grouse moors than elsewhere: 0.8 fledglings per breeding female per year, compared with 2.4 on other moorland and 1.4 in young conifer forests. Human interference with nest sites was documented on half of the grouse moor estates, accounting for at least 30% of breeding failures there. Moreover, annual survival of female hen harriers breeding on grouse moors was around half that on other moorlands, the cause being killing of the birds by humans. On average, 55–74 breeding females were killed each year over the eight-year study, representing 10–15% of the total population of breeding females in Britain (excluding Orkney).

Etheridge *et al.* further show that these losses on grouse moors would be unsustainable, leading to the rapid disappearance of hen harriers, were it not for breeding in the other habitats. Natal dispersal distances for both male and female hen harriers exceed 10 km, and harriers fledged from one kind of habitat were commonly found breeding in another. In short, moorland managed for grouse shooting is currently acting as a sink, with about 70% of its female recruits coming from other moorland or from conifer forests.

Overall, Etheridge *et al.* calculate that the total population of breeding female hen harriers in Britain is likely to be slowly falling. Their estimate is a decline of around 6% per year, but the many uncertainties are such that this is not significantly different from a zero rate of decline, or even a very small increase; conversely, the situation could be worse than estimated. We need a repeat of the Royal Society for the Protection of Birds' 1993 nationwide survey of hen harriers². But at the least Etheridge and colleagues' study suggests that, without the current persecution on grouse moors, Scotland's population of hen harriers would begin to increase at something like 13% a year.

From the gamekeepers' viewpoint, the indications are that grouse breeding success and survival would be higher in the absence

of hen harriers and other raptors^{5,6} (mainly peregrine falcons but to a small extent short-eared owls and golden eagles). So their persecution on grouse moors is not unreasonable (although it is against the law). We have a real problem here, with conservation interests on both sides: the grouse-shooting industry helps to conserve traditional British uplands and their vegetation, which would otherwise be likely to disappear under pine plantations; but the economic viability of this industry is seen by some of its practitioners to require the illegal killing and harassment of birds of prey. Underlying this dilemma, Etheridge *et al.*³ observe that "the sport shooting of red grouse and other game birds may not continue to be acceptable to the general public if its proponents argue that it can be sustained only by the persecution of rare birds of prey". The existence of an increasingly influential lobby against blood sports in general adds force to this argument.

So what else can be done? Phillips and Watson⁷, and others, have suggested that more active management of heather moorland, with optimal schedules of burning to

promote heather as food and cover for red grouse, could help. Although desirable, this approach would be expensive for a grouse-shooting industry that is arguably at the margins of profitability. Perhaps a more realistic way of reconciling the industry's interests with raptor conservation is to lay off the hen harriers and to concentrate more on the (legal) control of the relatively abundant crow and mammal predators of red grouse, particularly red foxes, whose "populations and distributions are less sensitive to culling than those of the hen harrier"³. □

Robert M. May is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. Newton, I. *Population Ecology of Raptors* (T. & A. D. Poyser, Berkhamsted, Herts, 1979).
2. Bibby, C. J. & Etheridge, B. *Bird Study* **40**, 1–11 (1993).
3. Etheridge, B., Summers, R. W. & Green, R. E. *J. Appl. Ecol.* **34**, 1081–1105 (1997).
4. Krebs, J. R. & May, R. M. *Nature* **343**, 310–311 (1990).
5. Redpath, S. M. *J. Appl. Ecol.* **28**, 659–671 (1991).
6. Hudson, P. J. *Grouse in Space and Time* (The Game Conservancy, Fordingbridge, Hants, 1992).
7. Phillips, J. & Watson, A. in *Heaths and Moorland: Cultural Landscapes* (eds Thompson, D. B. A., Hester, A. J. & Usher, M. B.) 344–361 (HMSO, Edinburgh, 1996).

Fluid dynamics

All stressed out

Mark Robbins

On page 360 of this issue¹, Peter Thompson and Sandra Troian report the discovery of remarkable behaviour at the interface between fluids and solids. They use a computer model to follow the motion of individual atoms in a fluid flowing over a solid surface. As the fluid is forced to flow more quickly, it suddenly begins to slip more easily over the solid. Even more surprisingly, the increase in ease of slipping follows a universal law. This transition in behaviour could affect macroscopic flow patterns in spreading, extrusion and other important industrial processes.

Scientists and engineers usually model fluid flow using continuum theories. Rather than following the motion of each atom, they describe the fluid by a few continuous variables that specify the average properties of atoms in each small region of space. These variables are the mean velocity, density and pressure. The Navier–Stokes equation relates these variables, and can be solved to see how they change with time.

The continuum approach works well if changes occur over lengths much greater than interatomic distances. It runs into trouble at the sharp interface between a fluid and a solid, where the type of atom changes abruptly. So such interfaces must be included as boundary conditions on the continuum equations. Historically, these boundary conditions were guessed at and then tested by solving the equations and comparing the

solutions with experiments. But tremendous advances in computing power now make it possible to test these boundary conditions with much greater precision and resolution than macroscopic experiments can².

Mechanical equilibrium imposes two simple constraints on interfaces in any steady, time-independent fluid flow: both the pressure perpendicular to the interface and the shear stress tangential to the interface must be the same on the two sides, or else the interface will accelerate. The remaining boundary condition is some relationship between the velocities on the two sides of the

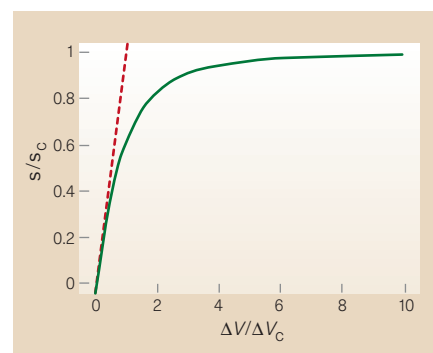


Figure 1 An apparently universal relationship between shear stress σ and the slip velocity ΔV between fluids and solids. The shear stress saturates at a value σ_c as the slip velocity diverges. ΔV is the slip velocity that would be observed at σ_c if the relationship were linear (dotted line).

interface. The usual, 'no-slip' boundary condition is an assumption that the velocities are the same (that is, fluid atoms don't slip over an adjacent solid).

This assumption describes most macroscopic flow patterns very well. The best-known exceptions are cases where a sharp corner forces the velocity to change rapidly, such as in the spreading of a fluid over a solid. The no-slip boundary condition would require the velocity within the fluid to change infinitely fast near the corner where the fluid interface intersects the solid, making it infinitely hard to spread butter or to get water to flow along a glass into your mouth. Fortunately, the no-slip boundary condition breaks down at molecular scales. In the past decade, molecular dynamics simulations^{3,4} have shown that large stresses can cause slip at the corner, and the flow here cannot be described by the Navier–Stokes equation.

Now Thompson and Troian have discovered a way for the no-slip rule to break down on macroscopic scales. They consider a fluid between two parallel plates. The bottom plate is stationary, and the top plate slides over it with velocity U at a fixed height h . If there is no slip, the velocity in the fluid goes continuously from zero (the velocity of the bottom wall) to U as the height increases from zero to h . The shear rate $\dot{\gamma}$ in the fluid, defined as the derivative of the velocity with respect to height, is then U/h . Any slip can be characterized by a length L_s which is the apparent increase in thickness of the fluid at each wall that would be needed to reconcile the velocity difference with the actual shear rate, $\dot{\gamma} = U/(h+2L_s)$.

Previous simulations have found many cases where the no-slip condition breaks down^{2,5}, but the slip length is usually only a few atomic diameters. This means that slip only becomes relevant in atomically thin films, where h is comparable to L_s . Thompson and Troian find similar results at low shear rates. However, as the shear rate increases towards a critical value $\dot{\gamma}_c$, they find a sharp divergence of L_s that follows a universal law. This means that L_s becomes comparable to the dimensions of ordinary pipes or nozzles, and would need to be included in models of fluid processing. One useful consequence should be a saturation in the friction of thin lubricating films.

Because $\dot{\gamma}_c$ has units of inverse time, it would be natural for it to reflect a characteristic timescale of atomic motion, such as the time taken for a fluid atom to respond to forces exerted by the solid. However, this is not born out by the simulations. Perhaps the transition is instead controlled by the shear stress σ . Within the fluids studied by Thompson and Troian, this is proportional to the shear rate. However, the stress is more directly associated with the interface, having the same value there as in the fluid, whereas the shear rate is only a property of the bulk

fluid. The divergence of L_s at $\dot{\gamma}_c$ implies that the velocity difference ΔV between the first layer of fluid and the solid diverges at $\sigma_c = \mu \dot{\gamma}_c$, where μ is the viscosity of the fluid. Equivalently, the maximum stress that the fluid–solid interface can sustain is σ_c . Figure 1 shows the functional relationship between σ and ΔV that is implied by Thompson and Troian's universal slip law.

The stress applied by the solid walls on adjacent fluid atoms comes from the variation of the potential energy of their interaction ϕ . Thompson and Troian find that $\dot{\gamma}_c$ and thus σ_c scales as a power of a 'roughness parameter' that is a function of the gradient

of ϕ . This is consistent with the idea that the transition is controlled by the shear stress. It will be interesting to see how surface roughness and other features of real surfaces affect Thompson and Troian's transition. □

Mark Robbins is in the Department of Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA.

1. Thompson, P. A. & Troian, S. M. *Nature* **389**, 360–362 (1997).
2. Koplik, J. & Banavar, J. R. *Annu. Rev. Fluid Mech.* **27**, 257–292 (1995).
3. Koplik, J., Banavar, J. R. & Willemsen, J. F. *Phys. Rev. Lett.* **60**, 1282–1285 (1988).
4. Thompson, P. A. & Robbins, M. O. *Phys. Lett.* **63**, 766–769 (1989); *Phys. World* **3**, No. 11, 35 (1990).
5. Thompson, P. A. & Robbins, M. O. *Phys. Rev. A* **41**, 6830 (1990).

Solvents

Molecular trees for green chemistry

Joan F. Brennecke

Because carbon dioxide is nontoxic, nonflammable, abundant and cheap, it ought to be every scientist's and engineer's favourite solvent for extractions, separations and reactions. Unfortunately, even at dense liquid or supercritical conditions, its ability to dissolve polymeric, ionic or highly polar species is exceedingly limited. On page 368 of this issue¹, Cooper *et al.* show how a fluorinated dendrimer can be used to extract strongly hydrophilic compounds from water into liquid CO₂, considerably expanding the applicability of CO₂ as a solvent.

CO₂ and 'green chemistry' have become inextricably intertwined in recent years (the latest example being the "1997 Green Chemistry and Engineering Conference" held in Washington DC on 23–25 June 1997). This may seem ironic in light of the bad press it has endured in connection with global warming, but using CO₂ as a solvent does not result in any net production. In fact, most CO₂-based processes are designed to recycle and reuse essentially all of the CO₂. With this in mind, the excitement about CO₂ has developed because of its potential to replace hazardous organic solvents, especially chlorinated liquids and freons. When leaked into the atmosphere these compounds contribute to stratospheric ozone

depletion, and the worst of them, the chlorofluorocarbons, have already been banned. In addition, many of the chlorinated solvents are carcinogens or suspected carcinogens, posing a health risk to chemical-plant workers, and to the general public if substantial leaks or spills occur.

CO₂ has other attractive features beyond its possible use as a substitute for less palatable organic solvents. Above its critical point (31 °C and 73.8 atmospheres), where the distinction between a liquid and a gas disappears, the density of CO₂ can be varied by almost an order of magnitude with relatively small changes in temperature or pressure — so its solvating power can be tuned and controlled. This should allow easy downstream separation: whether the dissolved species is a small organic molecule or a macromolecular assembly such as a micelle or a dendritic polymer, it can probably be enticed to separate from the CO₂-rich phase by appropriate adjustment of the solution density.

Supercritical fluids can also allow single-phase reactions, avoiding the limits imposed by mass transfer between two phases. In other reactions, rates and selectivities are better than can be achieved in liquids^{2,3}. Some of these attractive features have long been used commercially in the extraction of

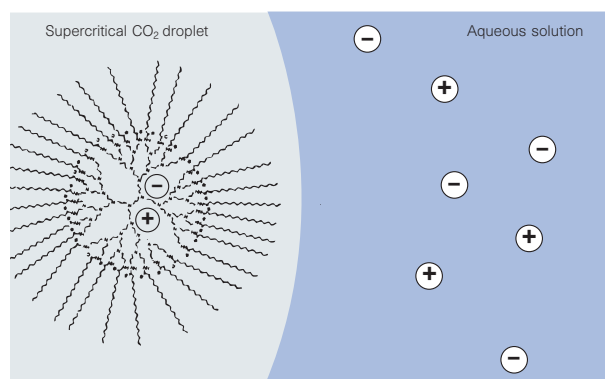


Figure 1 A dendrimer in a droplet of carbon dioxide. These highly branched polymeric surfactants have fluorinated tails, which make them soluble in non-polar CO₂, and they are able to pull into solution polar and ionic species that are normally insoluble in CO₂. This should help supercritical CO₂ replace environmentally damaging conventional solvents.

small, relatively nonpolar compounds⁴ (for example, the decaffeination of coffee and tea, and the extraction of flavours and aromas from natural products such as hops) and in the separation of similar compounds using supercritical-fluid chromatography. The two major drawbacks of supercritical CO₂, or liquid CO₂ at near-ambient temperatures, are the need for high pressures (approximately 50–350 atmospheres) and its difficulty in dissolving polymeric, ionic or highly polar species.

Whereas the requirement of high pressure is unavoidable, there have been a number of clever partial solutions to the limited-solubility challenge. An early idea was to add small amounts (1–5 mol%) of a co-solvent to the CO₂, generally intermediate in size and polarity between CO₂ and the target solute. This approach has been extremely successful and can increase the solubility of heavy organic solutes in CO₂ by an order of magnitude or more, especially where some specific chemical interaction such as hydrogen bonding occurs between the co-solvent and the solute.

This idea can be extended to extract metals into CO₂ by the addition of chelating agents. Conventional chelating agents such as acetylacetone and dithiones are effective, and their perfluorinated versions, as well as custom-made perfluoroether and silicone-based polymeric chelating agents, provide even greater solubilities of metals⁵.

For reactions that require both hydrophilic and hydrophobic constituents, one successful^{3,6} solution is the use of phase-transfer catalysis. But more recently, strategies for making hydrophilic compounds soluble have concentrated on forming organized molecular assemblies. A great success has been the identification of CO₂-loving perfluoropolyether surfactants, which in CO₂ form reverse micelles — hollow spheres of molecules with their hydrophilic heads on the inside. Their water-filled cores are capable of solubilizing proteins⁷ and supporting ionic chemistry⁸.

Now, in the work of Cooper *et al.*¹ we have an example of how a fluorinated dendritic polymer can be used in a similar way to further extend the applicability of CO₂ (Fig. 1). Dendrimers are highly branched polymers that form roughly spherical shapes in solution, acting like a micelle made out of a single surfactant molecule. Because the dendrimers used by Cooper and colleagues are fluorinated, they are soluble in CO₂, which is nonpolar. But the interior sections of the dendrimers are polar, and can trap highly polar or even ionic guest molecules. Thus the dissolved dendrimers make CO₂ a much better solvent for a variety of solutes.

CO₂, aided by 'smart' molecules, such as surfactants, dendritic polymers and chelating agents, will affect a wide range of industries. It can be used to replace freons or

methylene chloride (or other chlorinated liquids) in cleaning operations. It can remove metals and other hydrophilic contaminants from waste water, soils and sludges. It can be used for emulsion polymerizations. It might be able to extract a pharmaceutical product from a fermentation broth, or provide a 'microreactor' with a controlled intracellular environment for bioreactions. More importantly, all of these processes can be followed by a relatively easy separation step, which allows the reuse of an emerging environmentally friendly solvent. □

Joan F. Brennecke is in the Department of Chemical Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA.

1. Cooper, A. I. *et al.* *Nature* **389**, 368–371 (1997).
2. Roberts, C. B., Chateaufneuf, J. E. & Brennecke, J. F. *Am. Inst. Chem. Eng. J.* **41**, 1306–1318 (1995).
3. Morgenstern, D. A. *et al.* in *Supercritical Carbon Dioxide as a Substitute Solvent for Chemical Synthesis and Catalysis* (eds Anastas, P. T. & Williamson, T. C.) 132–151 (Am. Chem. Soc. Symp. Ser. 626, Washington DC, 1996).
4. Brennecke, J. F. *Chem. Ind. No. 21*, 831–834 (1996).
5. Yazdi, A. V. & Beckman, E. J. *Ind. Eng. Chem. Res.* **35**, 3644–3652 (1996).
6. Dillow, A. K. *et al.* *Ind. Eng. Chem. Res.* **35**, 1801–1806 (1996).
7. Johnston, K. P. *et al.* *Science* **271**, 624–626 (1996).
8. Clarke, M. J. *et al.* *J. Am. Chem. Soc.* **119**, 6399–6406 (1997).

Inhibitory synapses

Anaesthetics set their sites on ion channels

N. P. Franks and W. R. Lieb

The way in which volatile and gaseous general anaesthetics work has remained a puzzle ever since they were introduced into clinical practice during the 1840s. The prevailing view has been that anaesthetics (and, indeed, alcohol) do not have specific molecular targets in the conventional pharmacological sense, but that they nonspecifically disrupt nerve-membrane lipids^{1,2}. Evidence in support of this includes the molecular simplicity of the agents, and the correlation of potency with fat solubility.

Although this idea is, in some ways, an attractive unifying concept, it has fallen out of favour: anaesthetic effects on lipid bilayers are very small, and the potencies³ and stereoselectivities⁴ of general anaesthetics can be much more plausibly interpreted in terms of their binding directly to protein targets. Now, on page 385 of this issue, Mihic *et al.*⁵ report another step towards identifying specific molecular sites on ion channels, which may be involved in the actions of general anaesthetics and ethanol.

Chemical synapses are thought to be the

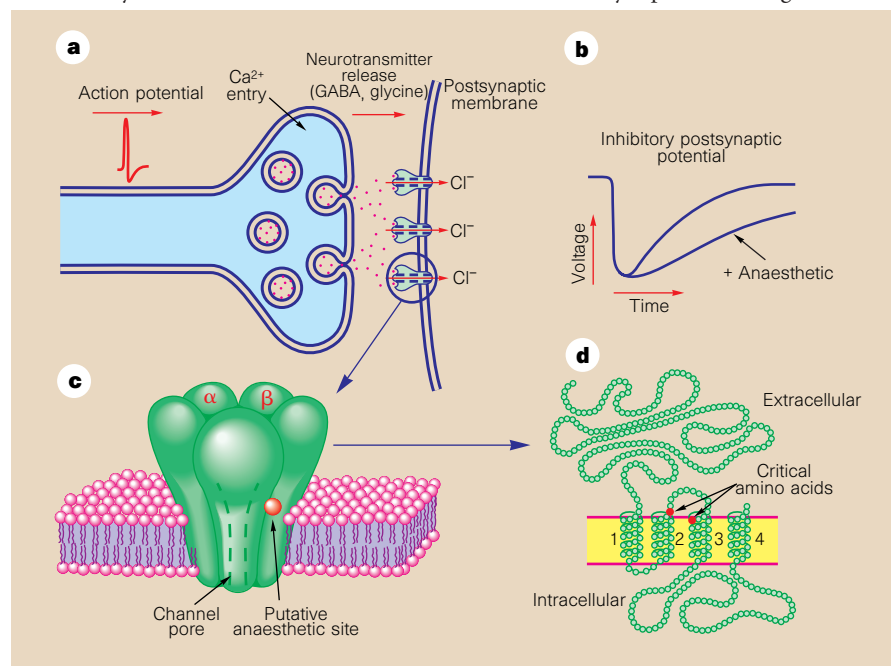
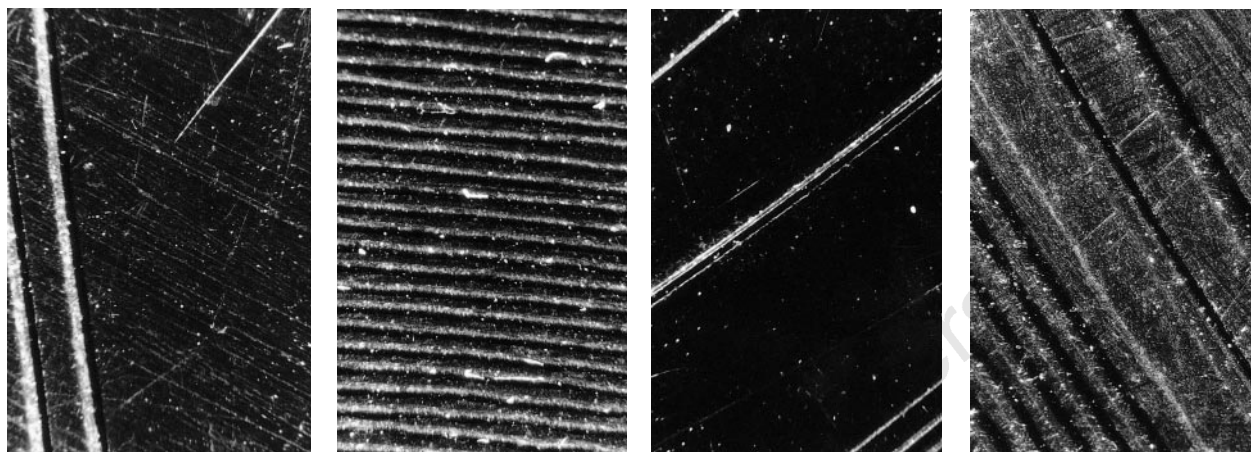


Figure 1 Mihic *et al.*⁵ have found that single amino-acid substitutions at two positions remove the potentiating effects of volatile anaesthetics and ethanol on GABA_A (γ-aminobutyric acid) and glycine receptors. a, GABA_A and glycine receptors bind the neurotransmitters that are released at inhibitory chemical synapses, and open to allow chloride ions to diffuse across the postsynaptic membrane. b, The main effect of volatile anaesthetics is to prolong channel opening and, hence, to increase postsynaptic inhibition. c, The receptor channels consist of pentamers of closely related subunits, and the structure of a single subunit is shown in d. The authors suggest that the two critical amino acids may form a binding site for general anaesthetics and ethanol.

Avoided objects

Nutcracker suite Grooves in a record that belonged to Hitler



main sites of general-anaesthetic action. At inhibitory synapses (Fig. 1), neurotransmitter — GABA (γ -aminobutyric acid) or glycine — is released from vesicles at presynaptic terminals in response to the voltage-dependent entry of Ca^{2+} , which is triggered by an action potential. The neurotransmitter diffuses to the postsynaptic membrane, and hyperpolarizes the membrane by opening inhibitory, chloride-permeable GABA_A or glycine receptor channels. These channels belong to an anaesthetic-sensitive superfamily of fast, neurotransmitter-gated receptor channels, which also includes the excitatory (Na^+ - and K^+ -permeable) 5-hydroxytryptamine (5-HT_3) and nicotinic acetylcholine (nACh) receptor channels⁶. Volatile anaesthetics inhibit the nACh receptors, but they markedly potentiate the activities of the other three types of receptor^{7,8}.

Using molecular genetics, binding sites for volatile anaesthetics that act on nACh receptors have been identified. Site-directed mutagenesis has revealed an inhibitory site within the aqueous pore of the muscle-type nACh receptor⁹. A second inhibitory site has been assigned to the large, neurotransmitter-binding, amino-terminal extracellular domain of another nACh receptor ($\alpha 7$). This was based on the behaviour of a homomeric receptor channel formed from chimaeric subunits — each subunit consisting of an amino-terminal domain from the nACh ($\alpha 7$) receptor, and a transmembrane and carboxy-terminal domain from the 5-HT_3 receptor¹⁰.

Mihic *et al.*⁵ have now identified a potentiating site near the extracellular regions of transmembrane domains 2 and 3 on the glycine and GABA_A receptors. They used an extensive set of well-designed chimaeras, consisting of complementary parts of a

glycine receptor and a $\text{GABA } \rho 1$ receptor. Also known as GABA_C , the $\text{GABA } \rho 1$ receptor is related to GABA_A , but it is inhibited — rather than potentiated — by volatile anaesthetics. From the behaviour of the chimaeric constructs, the authors deduced that a 45-amino-acid region was involved in potentiation by volatile anaesthetics and ethanol. By systematically mutating these amino acids, they identified two residues that were critical for anaesthetic potentiation. Interestingly, one of these also determines the potentiating effect of etomidate¹¹ (a clinically used intravenous general anaesthetic) on GABA_A receptors. But neither residue seems to be involved in the action of propofol, which is another intravenous agent.

Are these critical residues part of an anaesthetic-binding site, or are they involved only in gating or allosteric linkage? The answer is not known, and the issue is unlikely to be definitively resolved using molecular genetics. In some cases, such as the inhibitory pore site for anaesthetics on the muscle-type nACh receptor, an anaesthetic-binding site is probably involved, because the kinetics of channel inhibition are entirely consistent with the proposed site within the channel pore⁹. But in the case of potentiation — which is less well understood than inhibition — clear-cut interpretation is more problematic. Structural studies should help, although sufficiently high-resolution images of ion channels are still some way off. Yet only when the three-dimensional structures of anaesthetic-binding sites on channels are known can we even begin to think in terms of developing 'designer' general anaesthetics. These would bind specifically to selected ion channels, to produce general anaesthesia with fewer of the undesirable side-effects

that are associated with currently used agents.

Meanwhile, the use of molecular genetics to identify the amino acids involved in anaesthetic modulation of channel activities will be encouraged by the success of Mihic *et al.* The technique will undoubtedly be extended to other protein targets and classes of anaesthetic. But a broader problem remains. How do we extrapolate from specific sites on ion channels, at the molecular level, to the effects of anaesthetics on the intact animal? For example, how can we be sure that, even though the GABA_A receptor is sensitive to anaesthetics *in vitro*, these effects are essential to — or even contribute to — the state of general anaesthesia itself? Perhaps the most exciting aspect of these new findings is the prospect (suggested by Mihic and colleagues) of using transgenic-animal technology to express anaesthetic-insensitive mutant channels in mice. This would act as a selective test for the relevance and importance of various ion channels to the process of general anaesthesia. □

N. P. Franks and W. R. Lieb are in the Biophysics Section, The Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK.

1. Miller, K. W. & Pang, K.-Y. Y. *Nature* **263**, 253–255 (1976).
2. Lyon, R. C., McComb, J. A., Schreurs, J. & Goldstein, D. B. *J. Pharmacol. Exp. Ther.* **218**, 669–675 (1981).
3. Franks, N. P. & Lieb, W. R. *Nature* **310**, 599–601 (1984).
4. Franks, N. P. & Lieb, W. R. *Science* **254**, 427–430 (1991).
5. Mihic, S. J. *et al.* *Nature* **389**, 385–389 (1997).
6. Unwin, N. *Neuron* **10**, (Suppl.) 31–41 (1993).
7. Harris, R. A., Mihic, S. J., Dildy-Mayfield, J. E. & Machu, T. K. *FASEB J.* **9**, 1454–1462 (1995).
8. Franks, N. P. & Lieb, W. R. *Nature* **367**, 607–614 (1994).
9. Forman, S. A., Miller, K. W. & Yellen, G. *Mol. Pharmacol.* **48**, 574–581 (1995).
10. Zhang, L., Oz, M., Stewart, R. R., Peoples, R. W. & Weight, F. F. *Br. J. Pharmacol.* **120**, 353–355 (1997).
11. Belelli, D., Lambert, J. J., Peters, J. A., Wafford, K. & Whiting, P. J. *Proc. Natl Acad. Sci. USA* (in the press).

Geophysics

Probing Earth's dynamo

Peter Olson

Most of the planets in our Solar System, Earth among them, have magnetic fields that originate from self-sustaining dynamos; so, too, do the jovian satellites Io and Ganymede, according to the results from the Galileo mission¹. Planetary magnetism is a common phenomenon because only three basic ingredients are needed for a dynamo: a sufficiently large volume of electrically conducting fluid somewhere in the planet's interior, an energy source such as convection to circulate the fluid, and planetary rotation to organize the resulting fluid motions. That sounds simple enough. But the inner workings of planetary dynamos have remained stubbornly opaque, mainly because the requirement of a large volume of conducting fluid has prevented the construction of laboratory-scale dynamos and, until recently, has permitted only rudimentary numerical simulations.

On page 371 of this issue², Kuang and Bloxham report on numerical models of convection-driven dynamos that successfully reproduce many of the observed characteristics of the Earth's magnetic field. Perhaps the most fundamental result is that their model generates an external field dominated by an axial dipole component, like that of the Earth, with an intensity close to the present-day geomagnetic dipole moment. The Kuang–Bloxham dynamo also exhibits westward drift of the non-dipole magnetic

field, much like the historically documented westward drift of the geomagnetic field.

Publication of this geodynamo simulation follows on the heels of another successful numerical model by Glatzmaier and Roberts³. The two dynamos operate under broadly similar physical conditions in terms of convective energy, rotation rate and fluid properties, and they produce comparable external magnetic fields. It is only when they are sliced open to reveal their internal dynamics that important differences appear. It turns out that these differences hold the key to understanding the dynamics of the Earth's core. Fortunately, predictions can be made from the Kuang–Bloxham and the Glatzmaier–Roberts dynamos that allow them to be tested against each other.

Both of them indicate that the solid iron inner core, despite its relatively small size, plays a vital role in the dynamics of the core as a whole. First, it is an important energy source; crystallization of iron at the inner–outer core boundary releases both latent heat and light constituents, driving convection in the liquid outer core. Second, the inner core presents a mechanical barrier to flow in the liquid outer core. It is how the liquid responds to this barrier that leads to major differences between the dynamo models.

Barriers to flow in the atmosphere and ocean are common because of stable density

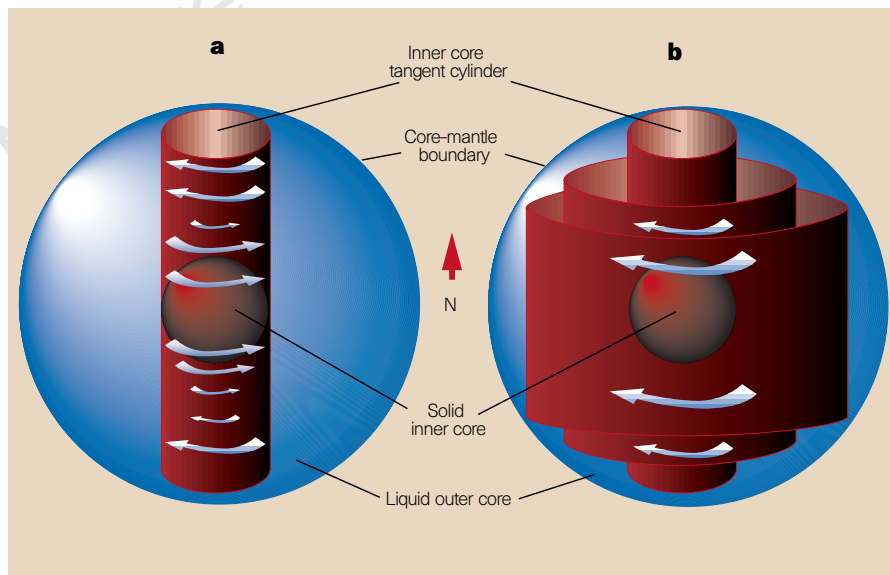


Figure 1 Alternative models of the primary azimuthal flow in the Earth's liquid outer core, which stem from setting different boundary conditions for the interaction between the inner and outer cores. a, Polar vortex flow inside the inner-core tangent cylinder, as surmised by Glatzmaier and Roberts³. b, Differential rotation outside the tangent cylinder, the result obtained by Kuang and Bloxham². Dynamo models with polar vortices account for the observed super-rotation of the solid inner core, whereas models with differential rotation outside the tangent cylinder account for the geomagnetic westward drift.



100 YEARS AGO

Professor Newcomb therefore concluded: "Either the bodies which compose our universe are vastly more massive and numerous than telescopic examination seems to indicate, or 1830 Groombridge is a runaway star, flying on a boundless course through infinite space with such momentum that the attraction of all the bodies of the universe can never stop it." A new contribution to this inquiry was read recently before the American Philosophical Society by Mr. Luigi D'Auria In this paper it is shown that, "given the ether the density as estimated by Maxwell, and the power of attracting matter by gravity, a body placed within the sphere of ether containing all the stars of the visible universe, and at a distance from the centre of such sphere equal to that passed over by light in 2200 years, would pass this centre with a velocity equal to that of the star 1830 Groombridge, taking into account the attraction of the ether alone; and such body would oscillate about the same centre, rectilinearly, with a period of a little over seven million years, which would be also the period of oscillation of every other star." Mr d'Auria recognises that some other, and unquestionable, cause may eventually prove to be responsible for stellar proper motions, nevertheless he thinks his results are worth putting on record.
From *Nature* 23 September 1897.

50 YEARS AGO

In 1681 Papin invented his "New Digester or Engine for Softening Bones", in which animal products are stewed by heat under pressure; this was a great advance in technique fraught with many valuable applications in subsequent years. To limit the pressure in the digester and so avoid bursting it, Papin employed the lever safety valve, with the invention of which he is not quite correctly credited; what is certain is that he was the first to illustrate it in print.... John Evelyn, in his "Diary" under date April 5, 1682, tells with gusto and appreciation of a supper cooked with the digester and partaken of in company with other fellows of the Royal Society after a meeting. He remarks: "This philosophical supper caus'd much mirth amongst us and exceedingly pleas'd all the company", as one can well imagine.
From *Nature* 27 September 1947.

stratification; examples include the atmospheric tropopause and the oceanic thermocline. In the Earth's core, convection tends to homogenize the fluid so the results of stratification are minor; instead, the effects stemming from rotation create a barrier. Rotation imparts a dynamical rigidity to a homogeneous fluid in the direction parallel to the rotation axis, through a principle known as the Taylor–Proudman effect. Rotation causes the fluid to circulate in coherent vortices as illustrated in Fig 1. Because the vortices resist stretching or bending, fluid in the outer core is inhibited from crossing an imaginary cylinder tangent to the equator of the solid inner core, the so-called tangent cylinder.

The main difference between the models comes in setting the boundary conditions between the inner and outer cores, which has dramatic effects on the tangent-cylinder barrier and on azimuthal flow in the outer core (Fig. 1a compared with Fig. 1b). When Kuang and Bloxham take the same rigid-wall boundary conditions on the inner core as used by Glatzmaier and Roberts, they obtain a similar dynamo — that is, one which features an intense polar vortex in each hemisphere inside the tangent cylinder and intense toroidal magnetic fields in that region. Through coupling with these polar vortices, the inner core is spun into super (prograde) rotation relative to the mantle and crust, as first predicted in the Glatzmaier–Roberts model. When Kuang and Bloxham change the boundary conditions to free-slip, in an attempt to reduce the effects of viscosity in the model and make the balance of forces more like the inferred conditions in the Earth, the polar vortices are largely suppressed and the toroidal field is induced by the azimuthal flow outside the tangent cylinder that gives rise to a westward geomagnetic drift. In this case, anomalous rotation of the inner core is weaker and features episodes of both retrograde and prograde relative motion.

Is there geophysical evidence for a barrier (6,000 km tall, according to theory) at the inner-core tangent cylinder? Surprisingly, there is. Most of the magnetic field that forms the Earth's exterior dipole field emerges from the core in flux patches located where the tangent cylinder intersects the core–mantle boundary⁴, just the location predicted by convective dynamo models. Even more dramatic evidence comes from the seismic detection of the super-rotation of the inner core^{5,6}, which preliminary investigations indicate rotates faster than the mantle and crust by 1–3° per year. This is close to the rate predicted by the dynamo models using rigid boundary conditions, a strong point in their favour. As more seismological data become available, however, the estimates of the anomalous inner-core rotation will be subject to revision,

so the final verdict is not yet in.

Realistic experiments on the mechanisms of planetary dynamos will remain out of reach for now, so numerical simulations provide the only way forward. Perhaps the most heartening outcome of the comparison of these two models^{2,3} is that it shows that such simulations are becoming increasingly powerful and accurate tools for relating magnetic fields to the dynamics in planetary interiors. □

Biopolymers

The structure of starch

Paul Calvert

Starch granules provide an insoluble but readily biodegradable storage system for plants. This balance of mechanical stability with degradability arises somehow from the chemical structure of the amylose and amylopectin that make the substance up, and from the arrangement of amorphous and crystalline zones within the granules. Starch also has considerable potential as a biodegradable plastic, but we need to understand the structure better before we can approach the ideal of a controllable service life followed by rapid degradation. A new high-resolution X-ray study by Waigh *et al.*¹ allows us for the first time to track across the structure of an individual granule.

Structural polysaccharides, such as starch, cellulose and chitin, undergo simultaneous polymerization and crystallization. This poses a structural constraint: how does the enzymatic machinery convert dissolved monomers into a solid polymer without getting choked by the solid as it forms? In chitin, little is known about how the solid is laid down. Cellulose, in contrast, has been much studied, and is extruded by an organelle that traces a spiral across the cell surface. Starch has the added constraint that it must lend itself to later controlled depolymerization, as it is used to fuel plant growth. It might be formed by a retreating enzymatic peripheral layer, or by entrained enzymes.

Because starch is such an important product, one would assume that the inner structure of granules had long since been analysed to death by electron microscopy. This has been done² but there are problems with artefacts, which have always dogged electron microscopy of polymers. The radiation intensity in a transmission electron microscope is so high that organic structures are usually destroyed in seconds. Hence there is always a concern about artefacts due to beam damage and, for starch, due to loss of water during preparation.

In the new study, pure and intense radiation from the ESRF Synchrotron in Grenoble was used in a 2- μ m-wide X-ray beam that could be scanned across single granules to

Peter Olson is in the Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland 21218, USA.
e-mail: olson@gibbs.eps.jhu.edu

1. Kivelson, M. G. *et al.* *Nature* **384**, 537–541 (1996).
2. Kuang, W. & Bloxham, J. *Nature* **389**, 371–374 (1997).
3. Glatzmaier, G. A. & Roberts, P. H. *Science* **274**, 1887–1890 (1996).
4. Bloxham, J. & Jackson, A. J. *Geophys. Res.* **97**, 19537–19563 (1992).
5. Song, X. D. & Richards, P. G. *Nature* **382**, 221–224 (1996).
6. Su, W.-J., Dziewonski, A. M. & Jeanloz, R. *Science* **274**, 1883–1885 (1996).

show the structure. It turns out that starch granules grow in rings, alternating between amorphous amylose and crystalline amylopectin (Fig. 1). The amylopectin crystals are built of double helices, which are arranged radially in the granule, so that the crystals are tangential to the surface.

In contrast, the spherulite structure found in most synthetic crystalline polymers has radial crystals, with the polymer chains running tangential to the surface. Such tangential chains make sense if they are thought of as adsorbing to the existing solid surface of a sphere. The radial chains seen in starch suggest, on the other hand, that monomer is being added to growing ends sticking out of the surface. So the structure of starch is determined by polymerization kinetics, not crystallization kinetics, which means that it will be difficult to arrive at an equivalent structure in a synthetic moulded part.

In natural polymers, the amorphous

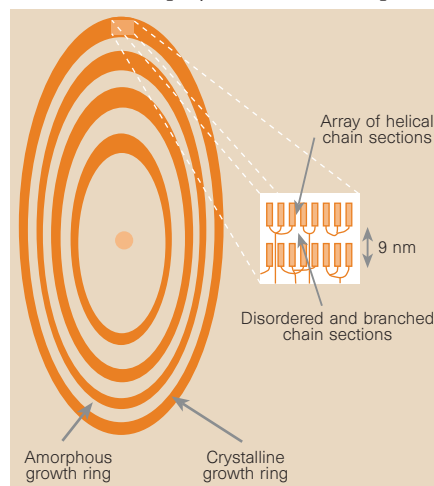


Figure 1 A schematic starch granule, as mapped by X-rays. Waigh *et al.*¹ see alternating shells of amorphous and crystalline material, with the amylopectin helices that form the crystal lying perpendicular to the shell. They also find that the granules are ellipsoidal rather than spherical, so the helices do not point to a single focus, implying that some elongated structure must initiate the growth.



Figure 2 One potato. Starches from wheat, banana and plantain are badly damaged by X-rays, so good diffraction patterns have not been obtained from them. But potato starch is tougher, probably because the crystalline layers are thicker, and so its structure can be determined.

regions are usually attributed to branched or irregular sections of the chains, which provide the organism with a means of controlling the stiffness and swelling of the polymer. Nuclear magnetic resonance studies of starch³ have detected amorphous regions which make up 70 per cent of the granule, some within the crystalline regions and the rest in distinct rings of highly branched and wholly amorphous material.

On the basis of electron microscopy, Oostergetel² has postulated a different model for potato starch, based on radiating double spirals of amorphous and crystalline material, like the swirls on a barber's pole, with an 8-nm cylindrical cavity down the centre of each spiral. So far, this model is consistent with, but is not really confirmed by, the new data, and it does seem to be very complex compared with what we might expect from simple polymerization and crystallization processes.

So what of starch as a basis for biodegradable plastic? A decade ago the development of bacterially derived polyhydroxybutyrate as a biodegradable thermoplastic was stalled by the high cost of fermentation and extraction, and that re-ignited interest in starch. Moulded starch golf tees and foamed packing material are already being made.

But the problem, as with most biological polymers, is that its properties are so sensitive to water. Dry starch will not melt but simply dehydrates, cross-links and then burns. Water plasticizes the polymer, allowing it to melt and be processed as a conventional plastic. However, the properties are very sensitive to water content⁴ and this is hard to control at moulding temperatures of 200 °C. In addition, hydrolytic degradation breaks the chains during moulding. As the plasticizing water is lost, a moulded piece of starch is

thus likely to shrink and become brittle. An alternative is to use glycerol in place of water, as it vaporizes less easily⁴, but it might leach out in use, again making the part brittle.

Natural starches differ in their amylopectin/amylose ratio and structure, probably to control the rate of hydrolysis. In reprocessed starch we have not attained a comparable level of control, but these new studies will teach us what kind of structures can be used to control degradation. It would also be valuable to transfer these lessons to biodegradable medical implants — normally made of polylactide or its copolymers with polyglycolide — in which it has proved very difficult to preserve mechanical properties past the first few per cent of the degradation reaction. We would really like a structure that retains its mechanical integrity to the point where much of the polymer has already been degraded and resorbed.

There is also renewed interest in grafting or blending starch with synthetic polymers to provide a combination of biodegradation with better mouldability⁵, but this would increase the cost. Instead, another substance — but one that is similar to starch — may be the answer: polyhydroxybutyrate moulds well and is not very water-sensitive. So perhaps what we are waiting for is a genetically modified corn that makes polyhydroxybutyrate instead of starch. □

Paul Calvert is in the Arizona Materials Laboratories, 4715 East Fort Lowell Road, Tucson, Arizona 85712, USA.

1. Waigh, T. A. *et al.* *Macromolecules* **30**, 3813–3820 (1997).
2. Oostergetel, G. T. & van Bruggen, E. F. J. *Carbohydrate Polymers* **21**, 7–12 (1993).
3. Morgan, K. R., Furneaux, R. H. & Larsen, N. G. *Carbohydrate Res.* **276**, 387–399 (1995).
4. van Soest, J. J. G. & Borger, D. B. J. *Appl. Polym. Sci.* **64**, 631–644 (1997).
5. Ramkumar, D. *et al.* *Eur. Polym. J.* **32**, 999–1010 (1996).

Daedalus

The light rolls in

Squash or deform a crystal of salt, and it will emit light. Many ionic crystals show this effect; the ionic reshuffling of deformation allows defects in the lattice to relax and release their stored energy. A crystal under deformation shines more brightly if it has been loaded with defects beforehand, for example by irradiation.

Daedalus wants to use the effect in a new lamp. He points out that the entire surface of an ionic crystal, at least in vacuum, is made up of defects. Imagine, he says, two ionic rollers rotating in contact, their surfaces being continually brought together and torn apart again. Their surface defects would be continuously healed and recreated. Copious light should be emitted, generated directly from mechanical power.

DREADCO physicists are now exploring the idea. The rollers should ideally be slightly plastic, so that they can deform readily and make intimate contact with each other. The most plastic ionic material is probably silver chloride. So the team is coating rollers with silver chloride and running them against each other in a vacuum, looking for light emission. The rollers are being irradiated and doped with various impurities, to generate excited surface states of the right energy for the colour of luminescence desired.

The 'Rolamp' should be far more efficient than existing lamps. It will be ideal for small hand-cranked torches, emergency beacons, bicycle lamps driven from the wheels, and so on. With little waste heat, it will run much cooler than filament lamps. Photographers and film and TV producers will welcome its coolness and economy, especially on location. Neater still, a Rolamp with an interrupted coating on one roller would emit light in brief flashes, as each active region made contact with the other roller. Synchronized with the camera shutter, it would emit light only when needed, using still less power and giving sharper images of moving objects.

The silver chloride could be doped to give light of any colour; a suitable mixture of dopants would give well-averaged white. But Daedalus likes the idea of distributing the different dopants around the rollers, so that this average is obtained from a rapid repeated sweep through the rainbow. The lamp would then excite our retinal colour sensors in a fast-cycling sequence. The slight time lags between our red, green and blue senses should give us a subtly heightened sense of colour separation.

David Jones

Obituary

John Cowdery Kendrew (1917–97)

Pioneer in structural biology and collaborative biological research in Europe

“... and in a jungle in Ceylon, J. D. Bernal talked with John Kendrew about the way it should be possible to use X-rays to solve the structure of proteins and so to understand their function in every living organism.” Thus wrote Dorothy Hodgkin about how wartime operational research, concerned with elephants and bombs, inspired Kendrew’s ensuing research career.

Sir John Kendrew died on 23 August, aged 80. He was educated at schools in Oxford and Bristol, and took his undergraduate degree (a first in chemistry) in 1939, at Trinity College, Cambridge. Like so many other scientists, during the Second World War he worked on radar, before being posted to southeast Asia as scientific adviser to the Allied Air Command.

At the outbreak of peace he sought out Max Perutz, an erstwhile student of Bernal’s. Impressed by Kendrew’s uniform as a wing commander, Perutz took him on as his second in command to work in the fledgling Medical Research Council unit for molecular biology at the Cavendish Laboratory, Cambridge, under Sir Lawrence Bragg (Kendrew may also have benefited from his entitlement to the residue of a senior scholarship from Trinity, which meant he didn’t cost too much). This extraordinarily fruitful collaboration led to the joint award of the Nobel prize for chemistry just over a decade later. Moreover, they recruited Francis Crick, Hugh Huxley, Jim Watson and Sydney Brenner to the unit, all of whom became prime movers in modern biology, and a dazzling array of post-doctoral students passed through the lab.

Perutz set out to solve the structure of haemoglobin, the oxygen carrier in blood. Kendrew chose myoglobin, haemoglobin’s small brother. The first attempt was with horse myoglobin but this would not crystallize properly. Try another species? Kendrew realized that myoglobin, an oxygen storage protein found in muscle, should be abundant in aquatic mammals. For reasons connected with the war, the nearby ‘low-temperature laboratory’ had cold rooms stuffed with whale meat, which duly yielded crystals of the quality necessary for X-ray diffraction.

How does one work out the structure of a molecule containing thousands of atoms? In principle it’s easy: you collect the X-ray diffraction data from the crystal in all



orientations and compute the Fourier transform to give the electron density. Unfortunately, the diffracted intensities yield only the amplitudes; the phases are lost, but it was thought that they might be recovered by adding a heavy metal at a specific site in each molecule. To everyone’s surprise this method worked and was used by Perutz to calculate a projection map of haemoglobin in 1953. Kendrew applied this method to myoglobin using three-dimensional data and many different heavy atoms, and in 1957 produced the first low-resolution map of a protein. It showed myoglobin to be almost entirely α -helical, the helices being coiled round each other and round the haem group.

Four years later Kendrew had the coordinates of the 1,250 non-hydrogen atoms in myoglobin, an epoch-making achievement. The enormous impact that these discoveries were to have on biochemistry was only slowly appreciated by classical biochemists, which in 1958 motivated Kendrew to found the *Journal of Molecular Biology*.

Kendrew ran his lab as chief of staff. He himself developed numerical techniques for finding the relative positions of the heavy atoms. Automatic methods for measuring the intensities of some 100,000 Bragg reflections were devised. Numerous computing ladies were entrusted with the collection and collation of all the data. Bigger and better X-ray sources were built. Early experiences had convinced Hugh Huxley that computing Fourier summations by hand was not a task befitting the human condition. As a result

he had introduced Kendrew to a friend, John Bennett from the Cambridge computer lab (EDSAC). In 1952, Bennett and Kendrew wrote a three-dimensional Fourier program, probably the first.

In 1962 Perutz and Kendrew were awarded the Nobel prize. That same year they moved into the splendid new MRC Laboratory of Molecular Biology on Hills Road, Cambridge, with Perutz as director, and Kendrew as deputy director and head of the division of structural studies.

However, he was restless, in search of new organizational peaks to conquer. In parallel to his Cambridge appointment he was deputy scientific adviser to the Ministry of Defence and later chairman of the Defence Scientific Advisory Committee, which earned him his knighthood. At the same time plans for a European Molecular Biology Laboratory, to be modelled on the particle-physics research facility CERN, were taking shape. Kendrew was appointed project leader and carried out all the intricate negotiations leading to the successful founding of this laboratory in Heidelberg in 1974 (he would have preferred Nice, but France had just got the new CERN ring so Germany must have EMBL).

In Heidelberg, he made the theme technology for molecular biology. He set up the EMBL outstations at DESY (Hamburg) and ILL (Grenoble) for synchrotron light as an X-ray source and for neutron scattering; the DESY outstation was the first in the world to use synchrotron radiation as a source for X-ray diffraction. In the Heidelberg laboratory he promoted many technologies, particularly cryo-electron and optical microscopy. Furthermore, he hired bright young project leaders and gave them freedom to develop. By the time he retired as director-general in 1982 he had built up a world-class laboratory.

Between 1982 and 1988 John Kendrew was vice-president and then president of the International Council of Scientific Unions, and he was a governor of the Weizmann Institute, Israel, where he had many friends. On leaving EMBL he was appointed president of St John’s College, Oxford, and was very happy to return to the city of his childhood and once more to work with students. In his Cambridge days he was a Fellow of Peterhouse College, and shortly before his death was awarded an honorary doctorate by the university, which occasioned him great pride.

K. C. Holmes

K. C. Holmes is in the Max-Planck-Institut für Medizinische Forschung, Abteilung Biophysik, Jahnstrasse 29, D-69120 Heidelberg, Germany. e-mail: holmes@otto.mpimf-heidelberg.mpg.de

A clupeid fish can detect ultrasound

It has been suggested that most teleost fishes cannot detect sounds higher than 2 or 3 kHz (ref. 1). However, we report here that at least one species of clupeid fish (herrings and shads), the American shad (*Alosa sapidissima*), can detect sounds up to 180 kHz. We speculate that clupeids are able to detect the ultrasonic clicks of one of their major predators, echolocating cetaceans².

Blueback herring swim away from echosounders, indicating that they may be able to detect ultrasound³, and ultrasound is used to keep several clupeids from entering water intakes of electric power-generating plants⁴. But it was not clear whether the fish were detecting the ultrasonic frequencies, or whether they were detecting lower-frequency side-bands associated with the pulsed sounds.

To test the range of frequencies detected, we measured auditory thresholds of shad (Fig. 1) using a classical conditioning technique in which the fish learned to reduce their heart rate when they detected a sound (Fig. 2a)⁵. We trained five fish using tones from 0.2 to 180 kHz (Fig. 2b). Measurement of the frequency spectrum of the signal showed that the sound contained only pure tones.

American shad showed the greatest sensitivity to sounds between 0.2 and 0.8 kHz (Fig. 2b). The thresholds were somewhat poorer at 1.6 kHz, with the worst sensitivity



Figure 1 A shoal of American shad, *Alosa sapidissima*, in their holding tank.

between 3.2 and 12.8 kHz. At ultrasonic frequencies (25–130 kHz), the thresholds were about 145 decibels (reference pressure 1 μ Pa), with a decrease in sensitivity at 180 kHz. Thus there are two regions of sensitivity for the American shad, one at low frequencies, as is commonly found among fishes, and one at high frequencies, at which most fishes have not previously been tested. The low-frequency thresholds were probably masked by background noise from pumps, and so cannot be considered absolute thresholds. Shad could also detect simulated bottlenose dolphin echolocation clicks, which had a 50- μ s duration, an inter-click interval of 50 ms and a peak frequency of 80 kHz.

To control for the possibility that the shad could have been detecting low-frequency or electrical artefacts rather than the ultrasound, we trained goldfish to detect a 0.8-kHz tone using the same procedures as for the shad. When the fish had learned to respond at a stable level we tested them with an 80-kHz tone. None of the four fish tested could detect this sound at the highest amplitudes that we could generate (160 dB, reference pressure 1 μ Pa), indicating that ultrasound detection by the shad was not artefactual.

Ultrasound detection is probably not widely found among fishes, although it has been shown that cod (*Gadus morhua*) can detect 38-kHz signals⁶. Our results demonstrate a far wider range of ultrasound detection in shad than in cod (which had a poorer threshold than the shad). But, as demonstrated by the goldfish, ultrasound detection is not necessarily common even among fishes that hear relatively well at low frequencies.

Clupeiformes have a unique ear structure in which a pair of thin air-filled tubes

project from the swimbladder and terminate in air chambers that are connected with the utricles of the inner ear⁷. This may allow the bullae and utricle to be used in ultrasonic hearing, although there is no evidence to show how this might take place.

The ability of clupeids to detect ultrasound may be an example of convergent evolution with moths and other insects which also have auditory systems capable of detecting the ultrasonic sounds of predators⁸. Shad readily detect echolocating pulses of dolphins, and, like moths, the response to the detection of such sounds is escape behaviour. It is possible that the ability of clupeids to detect ultrasound is a preadaptation that evolved before there were echolocating predators. All extant clupeids share the auditory specializations of the ear, and fossil clupeids are known from the Lower Cretaceous period (130 million years ago), long before odontocete cetaceans evolved in the Oligocene epoch (25–38 million years ago)^{9,10}.

**David A. Mann, Zhongmin Lu
Arthur N. Popper**

Department of Zoology, University of Maryland,
College Park, Maryland 20742-4415, USA
e-mail: Popper@zool.umd.edu

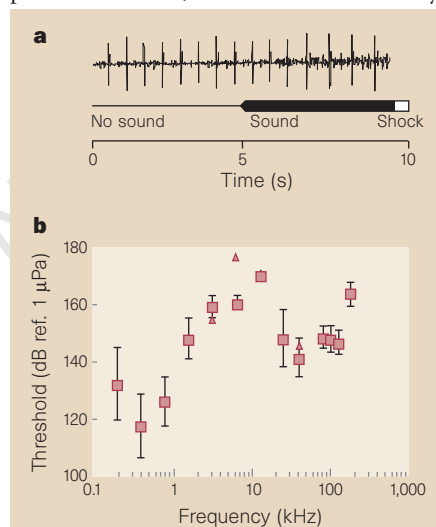


Figure 2 Ultrasonic detection by American shad. **a**, Example of cardiac responses to a 130-kHz sound followed immediately with a mild electric shock. After a 5-s control period with no sound, a tone was played for 5 s ending with an electrical shock. Individual spikes are heartbeats, which slow down in the presence of the sound and before the shock. **b**, Audiogram. Squares are means $\pm$ s.e.m., $N=5$. Triangles represent mean maximum amplitudes produced by the system at which some fish were unable to detect the sound.

1. Fay, R. R. *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Assoc., Winnetka, Illinois, 1988).
2. Au, W. W. L. *The Sonar of Dolphins* (Springer, New York, 1993).
3. Nestler, J. M., Ploskey, G. R., Pickens, J., Menezes, J. & Schilt, C. N. *Am. J. Fish. Man.* **12**, 667–683 (1992).
4. Ross, Q. E. & Dunning, D. J. *N. Am. J. Fish. Man.* **16**, 548–559 (1996).
5. Lu, Z., Popper, A. N. & Fay, R. R. *J. Comp. Physiol. A* **179**, 227–233 (1996).
6. Astrup, J. & Mohl, B. *J. Exp. Biol.* **182**, 71–80 (1993).
7. Blaxter, J. H. S., Denton, E. J. & Gray, J. A. B. in *Hearing and Sound Communication in Fishes* (eds Tavolga, W. N., Popper, A. N. & Fay, R. R.) 39–59 (Springer, New York, 1981).
8. Roeder, K. D. *Anim. Behav.* **10**, 300–304 (1962).
9. Nelson, J. S. *Fishes of the World* (Wiley, New York, 1994).
10. Barnes, L. G., Domning, D. P. & Ray, C. E. *Mar. Mam. Sci.* **1**, 15–53 (1985).

How accurate are satellite 'thermometers'?

We believe that lower-tropospheric temperatures measured directly by satellites have excellent long-term accuracy, as seen by comparisons with independent atmospheric measurements from weather balloons. Our results contradict indirect measurements by Hurrell and Trenberth¹ who claimed that the satellite data have significant discontinuities.

There are two independent direct methods currently used to measure the temperature of the atmospheric layer known as the lower troposphere (surface to 7 km). One method uses balloon-borne instrument packages (radiosondes) that rise through the atmosphere and the other uses satellite-measured intensities of microwave emissions from atmospheric oxygen, which are proportional to temperature (T_{2R} ; ref. 2). The latter method relies on polar-orbiting microwave sounding units.

Recently, Hurrell and Trenberth estimated the lower-tropospheric temperature indirectly using variations in sea surface temperatures (SSTs)¹. When comparing their SST-based estimates with T_{2R} in the tropics (20° S–20° N), they found two relative temperature jumps, one in September 1981 (0.25°C) and one in late 1991 (0.10°C). They claimed the jumps were due to problems with the satellite measurements.

To check the claim for 1981, we compared T_{2R} against temperatures measured independently at tropical radiosonde stations. We computed the average tropospheric temperature at each location for both radiosondes and satellites for the two years before and after September 1981. Using only four years of radiosonde data assures that instrumental inconsistencies are essentially non-existent. For stations in three latitude bands ($\pm 15^\circ$, $\pm 20^\circ$ and $\pm 25^\circ$ with 14, 30 and 38 locations, respectively) the two-year aver-

age temperatures after September 1981 minus those before are -0.04 , 0.00 and -0.02°C , respectively. The same differences for T_{2R} are -0.04 , -0.03 and -0.03°C . These two independent, direct measurements of tropical tropospheric temperatures before and after the alleged jump agree within $\pm 0.03^\circ\text{C}$, indicating no "spurious" jump of 0.25°C . Similar results are obtained for the "spurious" 1991 jump.

If these jumps described by Hurrell and Trenberth are real, trends over the period since 1979 should show clear differences of 0.1 to 0.2°C per decade between T_{2R} and radiosondes. Trends from several multi-station radiosonde data sets (one of which used over 300 stations worldwide and three data sets that had no changes in instrumentation) all agreed with T_{2R} to within 0.03°C per decade, the level of uncertainty we had previously ascribed to the satellite data^{3–6}.

Hurrell and Trenberth proposed the idea that merging NOAA-07 satellite data into the time series caused the "spurious" jump in late 1981. We disprove this claim by displaying the independent anomalies of tropical T_{2R} from NOAA-06 and NOAA-07 separately (Fig. 1). A displacement of 0.25°C should be clear, but we found nothing of that magnitude. The agreement between the satellites' anomalies, calculated directly, is 5 to 10 times better than Hurrell and Trenberth's approximation.

Additional speculations about the "spurious" jumps (surface emissivity variations or disagreements with the rigid SST/troposphere linkage found in climate models) are not supported by any observational evidence. Other studies have shown that surface emission variations do not significantly affect tropical averages (A. Basist, personal communication) and that multi-decadal variations in the surface/troposphere relationship do occur^{3,7,8}.

In summary, claims of large and spurious jumps in the T_{2R} satellite data set are not supported by either of two direct and independent measures of tropospheric temperatures.

John R. Christy
University of Alabama in Huntsville,
Huntsville, Alabama 35899, USA
e-mail: christy@atmos.uah.edu

Roy W. Spencer
NASA/Marshall Space Flight Center,
Huntsville, Alabama 35812, USA

William D. Braswell
Nichols Research Corporation/GHCC,
Huntsville, Alabama 35806, USA

Trenberth and Hurrell reply — It is difficult to create a continuous, consistent climate record from satellite observations alone. The main reason is that satellites and instruments have a finite lifetime and have to be replaced. For the current Microwave Sounding Unit (MSU) record, eight satel-

lites have been used. A significant problem is that each new satellite has a different orbit so that a different part of the diurnal cycle is sampled. This problem also afflicts individual satellites, as NOAA-6, NOAA-7 and NOAA-11 all drifted in equator-crossing time by between one and four hours.

Such drifts are important for MSU-2R, for which surface emissions account for nearly 20% of the signal over land, and are much less important for the middle tropospheric channel MSU-2. This is the reason for the lower noise (by a factor of three to five) in MSU-2 compared with MSU-2R, so that the reproducibility of temperature anomalies between different satellites is much better.

Surface emissions vary substantially with the diurnal cycle, so the observed orbital drifts mean that the diurnal cycle can get aliased with trends, especially over land. Such drift problems have been identified and corrected in the tropical outgoing longwave radiation (OLR) record from the same NOAA satellites: the NOAA-7 and NOAA-11 drift effects were especially large⁹. Because OLR also depends on cloud-top temperatures, the corrections do not translate directly to the MSU record, except perhaps over cloudless regions such as the Sahara where the greatest negative trends in the tropical MSU-2R record arise and, consequently, are suspect.

The discrepancy between the MSU-2R record and that implied by the surface record in the tropics shows up as two discontinuities¹, but it is likely to be more than that. Computing a single bias correction between two satellites that are drifting in orbit means that the statistics are not stationary, and a discontinuity can be introduced into the record in some locations, but other effects will also be present. This might be especially relevant to the NOAA-6 to NOAA-7 transition, where the magnitude of the discrepancy is larger than can be explained at present.

The precision claimed by Christy, Spencer and Braswell for MSU is overestimated. It is based on comparisons between overlapping satellite records but only after the offset bias and the annual cycle differences are removed. All of these are computed to minimize the differences and there is no validation using independent data. The evidence suggests that this procedure overfits the data and removes real variability from the records, thereby giving a false agreement between the residuals (their Fig. 1).

These satellite problems demonstrate that a climate observing system must also have an adequate ground-based *in situ* validation. For the MSU record, comparisons have been made with radiosondes, but measurement errors and sampling issues have a substantial impact on these records as well, especially over the tropical continents

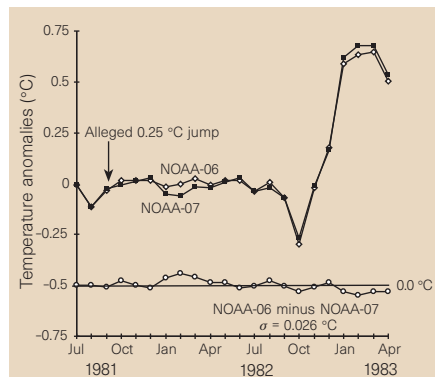


Figure 1 Monthly temperature anomalies ($^\circ\text{C}$) for the tropics (20° S–20° N) computed independently from two satellites, NOAA-06 and NOAA-07, for July 1981 to April 1983.

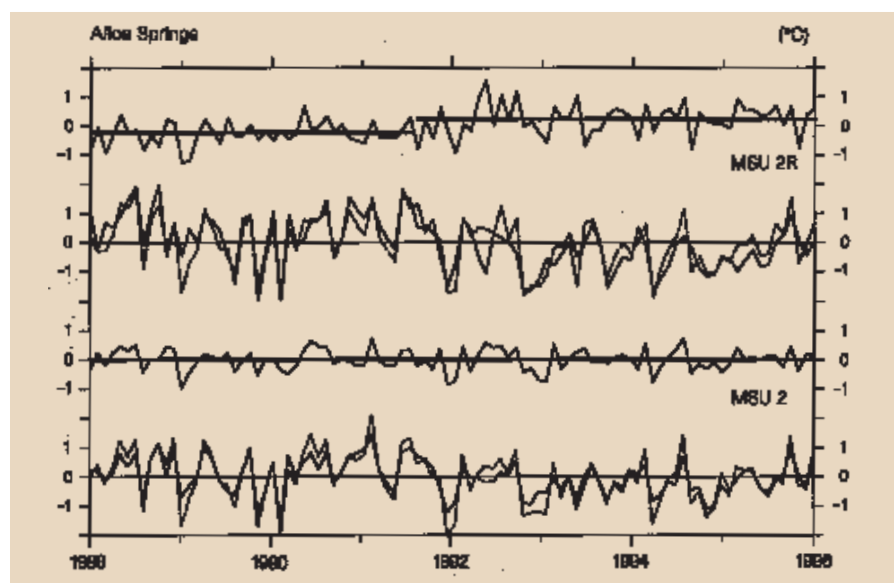


Figure 2 Comparison of temperatures from MSU and ground-based measures. Monthly mean temperature anomalies from MSU (solid lines) and Alice Springs (dashed lines), and the differences (Alice Springs–MSU), for MSU-2R (upper two curves) and MSU-2 (lower two curves). Mean differences are shown (heavy lines) after January 1988 for MSU-2, and before and after June 1991 for MSU-2R. Anomalies are relative to 1988–95 means.

where the cooling in MSU-2R is largest (-0.24°C per decade since 1979). Noise in MSU-2R over tropical oceans is a factor of two less than over tropical continents because of differences in surface emissions.

It is wrong, therefore, to assume that good agreement between a few, mainly tropical island stations validates the MSU products, and it is clear that such stations cannot adequately represent a larger-area average. A comparison with the radiosonde record at Alice Springs, Australia, after August 1987 (when a new radiosonde instrument was introduced), reveals a spurious stepwise cooling in the MSU-2R record not seen in the MSU-2 record (Fig. 2). This is consistent with the jump in the tropical MSU-2R record in mid-1991, which coincides with the NOAA-10 to NOAA-12 transition, identifiable in comparisons with tropically averaged sea surface temperatures¹.

Our argument is not that the surface and lower-tropospheric temperature trends must be identical, especially over short (17-yr) periods, but rather that the added noise in the MSU-2R record makes it unsuitable for trend analysis. Unfortunately, it is in this context that the data are most frequently referenced.

Kevin E. Trenberth, James W. Hurrell
National Center for Atmospheric Research,
Boulder, Colorado 80307, USA

- Hurrell, J. W. & Trenberth, K. E. *Nature* **386**, 164–167 (1997).
- Spencer, R. W. & Christy, J. R. *J. Clim.* **5**, 858–866 (1992).
- Angell, J. K. *J. Clim.* **1**, 1296–1313 (1988).
- Parker, D. E. *et al. Geophys. Res. Lett.* **24**, 1499–1502 (1997).
- Christy, J. R. *Clim. Change* **31**, 455–474 (1995).
- Christy, J. R., Spencer, R. W. & Lobl, E. *J. Clim.* (submitted).
- Reid, G. C., Gage, K. S. & McAfee, J. R. *J. Geophys. Res.* **94**, 14705–14716 (1989).
- Ross, R. J. *et al. Geophys. Res. Lett.* **23**, 3179–3182 (1996).
- Waliser, D. E. & Zhou, W. *J. Clim.* (in the press).

Frogs reabsorb glucose from urinary bladder

The amphibian urinary bladder is a bilobate, highly distensible and vascularized sac that stores fluid for use during periods of water stress¹. The organ is composed of a thin basement membrane overlaid by a selectively permeable ‘tight’ epithelium and is important in the homeostatic regulation of ion and osmolyte balance^{2,3}. We report here that glucose reabsorption from the urinary bladder permits recovery of sugar destined for excretion in the freeze-tolerant frog *Rana sylvatica*, whose unique winter survival strategy invokes extreme hyperglycaemia and ultimately, glucosuria.

We formulated the hypothesis that the urinary bladder might reabsorb glucose during studies of the wood frog, *R. sylvatica*, an anuran that is remarkably well-adapted to the cold, enduring the freezing of 65–70% of its body water during hibernation^{4,5}. Freeze tolerance in *R. sylvatica* is supported by physiological responses that mitigate cryoinjury, including a massive production of glucose⁶, which acts as a cryoprotectant⁷.

After thawing, frogs are extremely hyperglycaemic (up to 0.5 M glucose in the blood). Although some of the circulating cryoprotectant is directly reconverted to glycogen in the liver⁸, blood levels exceed the renal threshold for reabsorption and glucose is copiously excreted⁹. Glucosuria might critically limit the capacity to resynthesize cryoprotectant during subsequent freezing bouts, so we questioned whether *R. sylvatica* can effectively recover glucose

from bladder urine.

We studied glucose permeability using traditional solute-flux techniques with isolated hemibladders prepared from cold-acclimated *R. sylvatica*. We filled ligated hemibladders with a glucose-laden mucosal solution, formulated to approximate uretal urine of recently thawed frogs, and suspended them in hypertonic saline which we monitored for the subsequent appearance of glucose. The concentration of glucose in the serosal bath increased rapidly but ultimately approached the asymptote (Fig. 1). Net efflux from the bladder was strongly influenced by the initial glucose concentration of the mucosal solution (60 or 100 mM) and incubation temperature (4 or 23 °C).

To verify glucose uptake *in vivo*, we infused the bladders of live frogs with a glucose solution containing tritiated D-glucose. After incubating specimens for 4–6 h at 4 °C we scanned their tissues for the marker. Elevated radioactivity in the blood, lymph, and visceral and peripheral organs attested that the exogenous glucose had permeated the bladder epithelium and entered the general circulation. Control experiments confirmed that ligation of the bladder neck before administering glucose effected vascular isolation and confined the radiomarker within the bladder.

The concentration of glucose in urine produced by recently thawed frogs averages 80 mM (ref. 9), so reabsorption of this excreted sugar, even at the relatively low rates operative at winter temperatures (2–6 °C), seems crucial for restoring carbohydrate and energy balance during recovery from freezing.

Net efflux rates from isolated hemi-

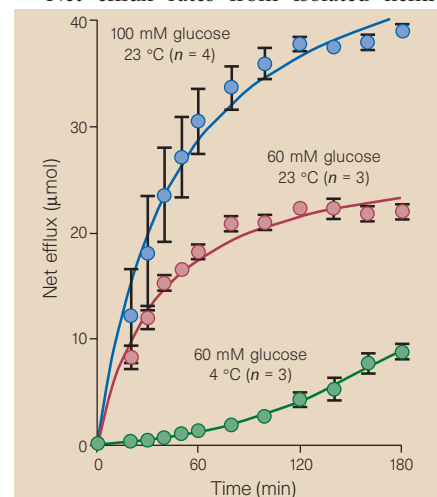


Figure 1 Time course for net glucose efflux. Efflux from isolated hemibladders of *R. sylvatica*, was based on incremental increases in glucose concentration of the serosal bath. Back flux was not determined. Hemibladders were inflated with 400 μl dilute phosphate-buffered saline (PBS; 35 mosM per kg) containing 60 or 100 mM glucose, and incubated for 180 min in a stirred serosal bath (PBS, initially 230 mosM per kg) at 4 or 23 °C. Means $\pm$ s.e.m.

bladder preparations were similar in *R. sylvatica* and the closely related, but freeze-intolerant, common frog (*R. temporaria*) of Europe and the leopard frog (*R. pipiens*) of North America. We also noted bladder permeability to glucose in the taxonomically distant bufonid, *Bufo marinus*, and the neotenic urodele, *Necturus maculosus*.

The taxonomic diversity of species exhibiting glucose permeability of the bladder indicates that this organ is fundamental for energy balance in amphibians whose carnivorous diet contains little carbohydrate¹⁰. The urinary bladder has long been used in studies of solute and water permeability, and may prove to be an ideal model for investigating transepithelial glucose flux.

Jon P. Costanzo, Phyllis A. Callahan

Richard E. Lee Jr, Michael F. Wright

Department of Zoology, Miami University,
Oxford, Ohio 45056, USA

e-mail: costanj@muohio.edu

- Bentley, P. J. *Science* **152**, 619–623 (1966).
- Boutillier, R. G., Stiffler, D. F. & Toews, D. P. in *Environmental Physiology of the Amphibians* (eds Feder, M. E. & Burggren, W. W.) 81–124 (Univ. Chicago Press, Chicago, Illinois, 1992).
- Shoemaker, V. H. & Nagy, K. A. *Annu. Rev. Physiol.* **39**, 449–471 (1977).
- Costanzo, J. P., Lee, R. E. Jr, DeVries, A. L., Wang, T. & Layne, J. R. *J. FASEB J.* **9**, 351–357 (1995).
- Storey, K. B. & Storey, J. M. *Annu. Rev. Ecol. Syst.* **27**, 365–386 (1996).
- Storey, K. B. & Storey, J. M. *J. Comp. Physiol.* **155**, 29–36 (1984).
- Costanzo, J. P., Lee, R. E. Jr & Lortz, P. H. *J. Exp. Biol.* **181**, 245–255 (1993).
- Russell, E. L. & Storey, K. B. *Cryo-Lett.* **16**, 263–266 (1995).
- Layne, J. R. Jr, Lee, R. E. Jr & Cutwa, M. J. *Herpetol.* **30**, 85–87 (1996).
- Pough, F. H. in *Behavioral Energetics: the Cost of Survival in Vertebrates* (eds Aspey, W. P. & Lustick, S. I.) 141–188 (Ohio State Univ. Press, Columbus, Ohio, 1983).

Parental age gap skews child sex ratio

The proportion of male to female births increases during and shortly after periods of war^{1,2}. We show that the age difference between parents (age of husband – age of wife) predicts the sex of the first child. We also find that in England and Wales, the mean spouse age difference increased during and immediately after the two World Wars and was strongly correlated with the sex ratio during the period 1911–52.

We obtained the age and sex of children from 301 families who attended secondary schools that recruited from a wide range of socioeconomic groups. The mean age difference D_a (age of husband – age of wife) was 2.48 years $\pm$ 0.23 (s.e.m.) and there were 301 first-born and 260 second-born children. Among first-borns there was an excess of daughters from couples with low D_a and an excess of sons from those with high D_a ($D_a = -9$ to -1 years: 14 sons and 29 daughters; $D_a = 0$ to 5 years: 117 sons

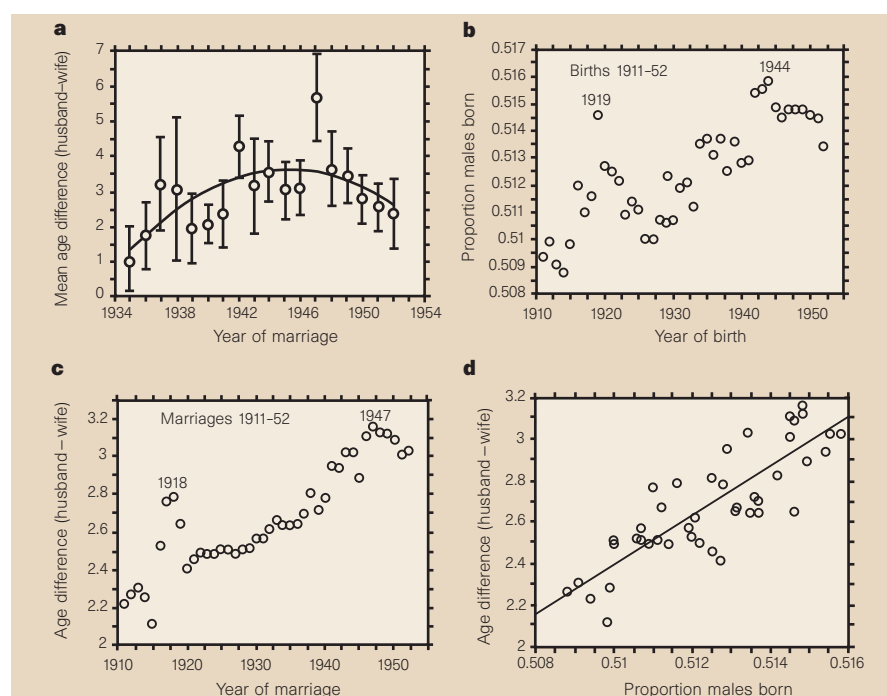


Figure 1 Parental age differences and sex-ratio statistics, 1911–52. **a**, The relationship between the mean ($\pm$ s.e.m.) of the difference in age between husbands and wives (D_a) and year of marriage (1935–52) in the Woolton area of Liverpool. There is a significant curvilinear relationship with a peak value of D_a in 1947 (second order polynomial, $y = -42.15 + 2.024x - 0.022x^2$, $F = 5.88$, $P = 0.013$, $n = 469$ marriages). **b**, Sex ratios of births registered in England and Wales from 1911–52; and **c**, D_a for marriages in the same period. **d**, Linear regression of sex ratio of births in England and Wales against D_a , 1911–52 ($r^2 = 0.68$, $F = 86.46$, $P = 0.0001$).

and 84 daughters; $D_a = 5$ to 15 years: 37 sons and 20 daughters; $\chi^2 = 11.86$, $P = 0.0027$). Among second-borns there was the opposite but non-significant tendency ($D_a = -9$ to -1 years: 22 sons and 11 daughters; $D_a = 0$ to 4 years: 93 sons and 89 daughters; $D_a = 5$ to 17 years: 20 sons and 25 daughters; $\chi^2 = 3.93$, $P = 0.14$).

The age of parents at the birth of the child has a weak effect on the child's sex³. However, multiple regression analyses with sex of child as the dependent variable and D_a and age of mother or father at birth as independent variables showed that D_a remained significantly associated with sex of child (D_a /age of mother – D_i : standardized partial regression coefficient $b_1 = -0.14$, $t = 2.35$, $P = 0.02$; age of mother: $b_2 = 0.13$, $t = 0.22$, $P = 0.83$; D_a /age of father – D_i : $b_1 = 0.14$, $t = 2.34$, $P = 0.02$; age of father: $b_2 = 0.13$, $t = 0.21$, $P = 0.83$).

Local and national patterns of D_a during the period 1911–52 (ref. 4) are shown in Fig. 1a, c. If couples do not delay the birth of their first child, D_a and sex ratio should be correlated and changes in the sex ratio should be preceded by changes in D_a . This is seen in 1914–18 but not during the Second World War (Fig. 1b, c). Registration of second and subsequent births will weaken the relationship between D_a and sex ratio so that an exact correlation is unlikely. Nevertheless a regression of sex ratio on D_a shows that the latter explains 68% of the variance of the former (Fig. 1d). Age of woman at

marriage was negatively related to the sex ratio ($b = -0.003$, $r^2 = 0.23$, $F = 12.19$, $P = 0.001$). However a multiple regression analysis with sex ratio as the dependent variable and D_a and bride's age as independent variables left D_a as the only significant correlate of sex ratio (D_i : $b_1 = 0.78$, $t = 8.26$, $P = 0.0001$; age of bride: $b_2 = -0.14$, $t = 1.51$, $P = 0.14$).

Rank in many animals is related to the sex of their offspring⁵. In humans, the elite often form partnerships with high D_a ⁶ and have more sons than daughters⁷. It may be that during wartime women prefer to marry older men with high resources and this leads to an increase in D_a . We do not know how the sex of first-borns is adjusted in relation to D_a . Women could influence the motility of sperm bearing either X or Y chromosomes or they may invest differentially in males and females *in utero* leading to higher miscarriage rates of one or the other sex.

J. T. Manning, R. H. Anderton, M. Shutt

Population Biology Research Group,

School of Biological Sciences,

University of Liverpool, Liverpool L69 3BX, UK

e-mail: jtmann@liv.ac.uk

- Martin, W. J. *Lancet* **1**, 807 (1943).
- MacMahon, B. & Pugh, T. F. *J. Hum. Genet.* **6**, 284–292 (1954).
- Bromwich, P. *Prog. Obstet. Gynaecol.* **7**, 217–231 (1989).
- The Registrar General's Statistical Review of England and Wales, Part II Civil (HMSO, London, 1921–1952).
- Clutton-Brock, T. H. & Iason, G. R. *Q. Rev. Biol.* **61**, 339–374 (1986).
- Kenrick, D. T. & Keefe, R. C. *Behav. Brain Sci.* **15**, 75–133 (1992).
- Mueller, U. *Nature* **363**, 490 (1993).

Tapping a rich vein of medical history

Drawing Blood: Technology and Disease Identity in 20th-Century America

by Keith Wailoo

Johns Hopkins University Press: 1997.

Pp. 288. \$39.95, £33

David Weatherall

In a self-revealing preface, Keith Wailoo describes the event that set him thinking about the sociological aspects of medical technology and which led eventually to this book. His father was admitted to hospital for investigation of kidney stones and during "a routine work-up" was found to have a non-specific abnormality of his electrocardiogram. As so often happens, expert cardiologist opinion was ambiguous, much anxiety was engendered and the significance of the finding was never explained. This all too common event led him to consider the ways in which the technologies of modern medicine have been used to classify sick people, and their broader influence on the emergence of new medical specialties, patient groups, institutions and society.

Surprisingly, therefore, Wailoo has chosen technical advances in haematology rather than cardiology to address these issues. He begins his investigation with the mysterious illness chlorosis, a disease that has now disappeared from the scene but which was frequently diagnosed in young women earlier this century; its name refers to the greenish hue to the skin, which, together with extreme lethargy and depression, characterized the syndrome. With the development of methods for examining the blood, it became clear that many of these patients had a hypochromic anaemia, which, it was later found, may indicate iron deficiency. These labels do not, as the new technocrats of haematology apparently suggested, say anything definitive about the genesis of chlorosis; both hypochromic anaemia and iron deficiency have many causes. Later writers thought that the disease may have reflected the social strains, diets and precarious state of iron balance at a particular stage of a young woman's development, an interpretation that the author concludes "highlights medicine's enduring desire to define, through technology, a legitimate female identity".

After describing how the brief reign of 'splenic anaemia' brought much needed credibility to abdominal surgeons, and the complex interplay of haematology, society and industry that evolved as more was learnt about drug-induced aplastic anaemias and pernicious anaemia, Wailoo then moves into the era of the new genetics and the problems of sickle-cell anaemia. Along the way he describes the way in which the rapid flowering

of haematology and the technological advances on which it was based were important in the development of new professional attitudes and institutions.

Wailoo's interpretation of the social effects of the discovery of sickle-cell anaemia is particularly interesting. The first slide-tests for sickling failed to distinguish the trait from the disease; even after Linus Pauling's discovery of sickle-cell haemoglobin and the elucidation of its genetic transmission, this confusion continued. In the 1970s, sickling became an important issue in the United States, and the ill-organized screening programme that followed led to discrimination against sicklers. Wailoo does not believe that these issues were put into perspective even when it was realized that the disease is not unique to black races and that its high frequency is, in fact, the result of a balanced polymorphism due to heterozygote advantage against malaria. He concludes that the technology that clarified the cause of this disease provided a "substrate for building compelling but myopic narratives about race relations". He points out that the discovery of sickle-cell haemoglobin was only the beginning of the era of molecular medicine and that, more recently, technologies developed to study the human genome, particularly in the field of predictive genetics, are generating completely new sets of doubts and uncertainties for society.

It is difficult to argue with Wailoo's basic thesis. Questions of health and disease have always had to be viewed in their broader social context; as scientific medicine evolved, it was inevitable that its technologies would have wide-ranging effects. And because medical research evolves slowly, often with only a partial understanding of the meaning of each

step, there is the constant danger that new disease labels and technologies may be applied before their true significance is appreciated; cancer genetics is a current good example. On the other hand, if medical historians are to help us learn from these mistakes, it is important that they examine every angle of the problem dispassionately and avoid a pejorative approach to these difficult issues.

Although the problem of chlorosis cannot be interpreted entirely in the light of the reductionist technology of modern haematology, and may never be, the social effects of the discovery of sickle-cell haemoglobin are more accessible. During the years of confusion about the 'trait' and the 'disease', it was inevitable that some false inferences were drawn about the biological relevance of sickling. Once the mode of transmission and population genetics of sickle-cell anaemia were understood, there should have been no excuse for this type of thinking, although, mainly through scientific illiteracy, it still persists. But the real basis for the disaster of the US sickle-cell screening programme in the 1970s was that it was rushed through without the support of a well-informed counselling programme. The reasons are extremely complex, reflecting political and lay group pressures, the 'rediscovery' of a neglected disease in a racial minority and the social climate in the United States in the 1970s. History is repeating itself today, with similar pressures to develop genetic screening programmes before the scientific knowledge on which they must be based has been adequately researched.

One of the problems (and fascinations) of Wailoo's book is that it is based entirely on the experience of the United States. There, anything new is taken up with immediacy and



Demons and blood-letting as depicted in an early-nineteenth-century medical text. From *A Quest for the Code of Life*, in which Liz Fletcher charts the transition of Hinxton Hall, near Cambridge, England, from a run-down country estate to the home of the new £180 million Wellcome Trust Genome Campus. Fletcher also chronicles the life and work of the trust's founder, Sir Henry Wellcome, and explores the science behind the Human Genome Project, while Roy Porter provides a historical perspective to the new genetics. The Wellcome Trust, £5 (pbk).

enthusiasm; where else is one's blood cholesterol level a popular topic for dinner-party conversation, or is one accosted by two bearded joggers while sitting in a park, as I was recently, and told that I was polluting New York City by innocently puffing my pipe? Undoubtedly an over-enthusiastic reaction to medical and technological advance, sometimes based on limited knowledge of the science involved, which may itself be at an early stage of development, can have unpredicted effects, as evidenced by the premature sickle-cell screening programme.

The story Wailoo tells is therefore important. Medical advance will always have a long phase of gestation. At each stage there will be a tendency for doctors or scientists to exaggerate or misinterpret what they have discovered, or for politicians, companies or lay pressure groups to wish to use this information prematurely or inappropriately. The effects can be wide ranging and not always for the good of society. Social historians of medicine therefore have an important part to play in trying to analyse why this happened in the past and how these problems might be avoided in the future. For these reasons, even though the book's tone is pejorative in parts, I recommend it as a thought-provoking view of how new medical technologies can shape professions and society, not always to their benefit. The author has certainly irritated this clinician by some of his writing; but perhaps that is precisely what social historians should be doing if their work is to have the influence it deserves. □

Sir David Weatherall is at the Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Killing and culture

The Chemical Weapons Taboo

by Richard M. Price

Cornell University Press: 1997. Pp. 233.

\$45.25, £25

Henrietta Wilson

One of the more intriguing aspects of the history of war technology is the position of chemical weapons. As we know them today, chemical weapons were developed at the turn of this century, but they have been used in few conflicts, often despite their ready availability in belligerents' stockpiles. The First World War saw the first systematic use of chemical warfare, which caused 1.3 million casualties, and the first deployment of both tanks and aeroplanes. Unlike tanks and aeroplanes, however, which have become the mainstays of conventional warfare, chemical weapons have never been assimilated into military culture. Chemical warfare was used only six times during the interwar years, and its non-use during the Second World War is frequently noted with surprise, because the protago-



An ill wind: German soldiers discharge poison gas from cylinders in the First World War.

nists had far greater stockpiles of chemical munitions than the total quantity consumed in the First World War. Since the Second World War, chemical weapons have been used in only a handful of wars. This clearly contrasts with the escalating use of planes after the First World War, for example.

Such observations give rise to the notion that there is a taboo against the use of chemical weapons, setting them apart from other weapons. This aversion to using chemical weapons has been codified in international laws for some time. Modern chemical weapons were prohibited by the Hague Declaration of 1899, in which the signatories agreed "to abstain from the use of projectiles the object of which is the diffusion of asphyxiating or deleterious gases". After the First World War there was renewed commitment to devising effective legislation against chemical weapons, culminating in the 1925 Geneva Protocol. Most recently, chemical weapons have become the subject of the most comprehensive multilateral arms control regime so far, the 1993 Chemical Weapons Convention. This came into force in April 1997, and prohibits the development, production, stockpiling and use of chemical weapons.

In *The Chemical Weapons Taboo*, Richard Price elucidates the nature of the norm against the use of chemical weapons, studying how they have been singled out as particularly reprehensible. He bases his analysis on the "genealogical method" of Nietzsche and Foucault, describing this as an investigation into the discourses that reveal the origin and meaning of particular categories. He explores how chemical weapons have been thought about and represented, and what this tells us about the chemical weapons taboo. He considers the way chemical weapons have been treated in successive

international legislations, national policies and popular journalism, from the Hague Declaration to the 1993 convention.

The book is a valuable contribution to a difficult and unusual subject. Price emphasizes that there is nothing obvious or straightforward about the moral stigma attached to chemical weapons, revealing how responses to chemical weapons have been neither objective nor consistent. An example is the different ways in which the use of chemical weapons have been justified. In the First World War, the deployment of chemical weapons was often accompanied by claims that they were being used only in retaliation, or that they were no more inhumane than other weapons. By contrast, Price recounts that, far from arguing that the chemical weapons it used against Ethiopia in 1936 were humane, Italy did not even admit that it had used them until 1996. This difference in attitudes does not reflect any significant change in the type of weapons used.

The study has a broad and ambitious scope, covering much material. But there is some carelessness in the attributions. For example, Robert Harris and Jeremy Paxman are cited as among "the authors of the most systematic analyses of World War II", whereas their book *A Higher Form of Killing* is in fact a popular and reliable history of chemical and biological warfare throughout the twentieth century.

What is original about *The Chemical Weapons Taboo* is its exclusive emphasis on discourses, dictated by its use of the genealogical method. This is interesting, but potentially misleading. A typical criticism of the method is that discourses can be presented selectively, inviting accusations of incomplete, and possibly prejudiced, accounts. In this case, the selection of particular discourses is left unjustified and their context is

not discussed. The implication is that we are being given an exhaustive list of representations of chemical weapons. But there are many omissions, major and minor. Particularly notable is the neglect of military attitudes, important in understanding why the military have been especially averse to chemical weapons. An important question here is why chemical weapons have never been assimilated into core military doctrine or practice. Generally, weapons innovations are either accepted and incorporated into standard military theory or abandoned altogether for reasons such as military utility or institutional resistance. Chemical weapons are unusual in that, although not accepted into mainstream military practices, and despite the fact that they have rarely been used, their development and acquisition have often been maintained.

Another omission is any analysis of scientific discourses, which would outline how the science of chemical weapons has been defined and developed, and indicate the particular innovation processes involved. Scientists have often been the stimulus for the development and use of chemical weapons. For example, Fritz Haber, father of nitrogen fixation, championed the first use of chemical weapons in the First World War.

The 1993 Chemical Weapons Convention could in the coming decades be the frontrunner in a growing body of arms-limiting legislation. Price's book may contribute to this process, demonstrating how weapons systems can be made political and how taboos may be, in part, a cultural construct. □

Henrietta Wilson is at the Pugwash Conferences on Science and World Affairs (British Group), Ground Floor Flat, 63A Great Russell Street, London WC1B 3BJ, UK.

Let it shine

The Fire Within the Eye: A Historical Essay on the Nature and Meaning of Light

by David Park

Princeton University Press: 1997. Pp. 377.

\$29.95, £24.95

William Shea

One of the triumphs of the scientific revolution of the seventeenth century was the understanding of optical phenomena. René Descartes explained how the rainbow was produced by reflection and refraction in drops of water in the atmosphere, and Newton became famous for showing that white light is a mixture of different colours. Newton was celebrated in verse, and the admiration of his contemporaries is captured in Pope's famous epitaph: "Nature and Nature's laws lay hid in night, / God said, 'Let Newton be!' and all was light".

Nearly a century later, a group of poets

met in London in the studio of the painter Benjamin Haydon on 28 December 1817. In his diary, Haydon tells us how Wordsworth, "in a strain of humour beyond description, abused me for putting Newton's head into my picture: 'a fellow,' said he, 'who believed nothing unless it was as clear as the three sides of a triangle'. And then Keats agreed that he [Newton] had destroyed all the poetry of the rainbow by reducing it to its prismatic colours. It was impossible to resist him, and we all drank 'Newton's health, and confusion to mathematics'."

Not long after this toast, Keats wrote the familiar lines in *Lamia*: "Do not all charms fly / At the mere touch of cold philosophy? / There was an awful rainbow once in heaven: / We know her woof, her texture; she is given / In the dull catalogue of common things. / Philosophy will clip an Angel's wings, / Conquer all mysteries by rule and line, / Empty the haunted air, and gnomed mine / Unweave a rainbow...".

The wheel of fortune had come full circle. Worshipped, almost deified, by the Augustan poets for his successful explanation of the nature of light, Newton was now damned for destroying its numinous character. This incident is a powerful reminder that light is more than something we perceive with our eyes or measure with our instruments. When the Bible says, "God is light and in him is no darkness at all", we see a metaphor, but David Park argues in this genuinely illuminating book that the author of

these words may have intended them as a plain statement of fact. There were times when understanding and vision, or goodness and light, were so close that you could not slip a straw between them. But the history of thought resembles the astronomer's expanding Universe. As our mental cosmos widens, ideas once closely related spread out and become isolated and lose all connection, except, perhaps, what is remembered in a few words. Medicine once went hand in hand with astrology; little now remains of their relation except the term 'influenza', which refers to heavenly influences.

Writing for a general audience, Park avoids technical language and concentrates on thinking more than on results. He traces the development of our notions about the nature of light in a way that will interest those who know nothing about lasers as well as those who use them every day. The quest for the nature of light began with the study of vision: how does it happen that when we open our eyes our minds are filled with images of things close and far away? Why must there be light before we can see? For a thousand years people debated what the visual message is: what is it that carries the message, and does it move from the object to the eye or issue from the eye to play on the object? As late as the seventeenth century, as distinguished a natural philosopher as Descartes still believed that exceptionally gifted people (including himself) could see in the dark because their eyes



The ten chromosomes of corn

Mutants of Maize by M. G. Neuffer, E. H. Coe and S. R. Wessler contains more than 400 colour photographs of the diverse genetic variants of corn. First published in 1968 and long out of print, it now covers twice the number of mutants and provides detailed descriptions of mutant

gene loci as well as summaries of the physical structure of cloned genes. Mutant maize genes have counterparts in most higher plant species, so the book should have a wider readership than just corn biologists. Cold Spring Harbor Laboratory Press, \$250 (hbk), \$100 (pbk).

acted as searchlights. As a confirmation of this theory, he appealed to the overwhelming evidence that cats can see at night!

For what, indeed, is this thing we call light? It brings life to the world around us, and it is the means by which most of us see and act. At the end of the twentieth century, physicists call light a quantized electromagnetic field. But what is that? We understand the mathematical theory very well, but how to translate its rarefied jargon into the rich and specific language of the everyday world of common sense? A quantized field has a mathematical description in terms of waves or in terms of particles. It is a matter of convenience, for the two descriptions are in every way equivalent. If we want to talk about it with someone not interested in calculation, what should we say? In plain words, what is actually light? This is the story of how scientists have tried to answer this question. It is a fascinating account, brilliantly told, full of digressions that are useful because they help us to enter the intellectual and social climate where ideas about light were born. It is also a story with a moral: we come from darkness and end in darkness, but we must never stop saying, "Lead on, kindly light". □

William Shea is at the Institut d'Histoire des Sciences, Université Louis Pasteur de Strasbourg 7, rue de l'Université, 67000 Strasbourg, France.

Also on history

Einstein: A Life by Denis Brian. A well-researched biography first published in hardback last year. Wiley, \$19.95, £11.99 (pbk).

Galileo on the World Systems edited by Maurice A. Finocchiaro. A new abridged translation and guide. University of California Press, \$19.95 (pbk).

Equivalence and Priority: Newton Versus Leibniz by Domenico Bertoloni Meli. Includes Leibniz's unpublished manuscripts on Newton's *Principia*. OUP, £25 (pbk). Reviewed by W. R. Shea in *Nature* 364, 681 (1993).

Victorian Science in Context edited by Bernard Lightman. Wide-ranging essays, from Maxwellian physics to science fiction to zoological taxonomy, that look at how science influenced and was influenced by the larger Victorian culture. University of Chicago Press, \$70, £55.95 (hbk); \$22.50, £17.95 (pbk).

Discovering Birds: The Emergence of Ornithology as a Scientific Discipline, 1760–1850 by Paul Lawrence Farber. First published in 1982, the book begins by looking at the influence of Buffon's *Histoire naturelle*. Johns Hopkins University Press, £12.50 (pbk).

A Chronology of Medicine and Related Science by Leslie T. Morton and Robert J. Moore. Entries cover historical events; the establishment of institutions, hospitals, societies and journals; the publication of important books and papers; and biographical information for individuals. Scholar/Ashgate, £75

Truth on the outside

On the Surface of Things: Images of the Extraordinary in Science

by Felice Frankel and George M. Whitesides
Chronicle Books: 1997. Pp. 160. \$35 (hbk), \$22.95 (pbk)

Ronald Hoffmann

On reading this book one gets the immediate feeling that one's eyes and mind have truly feasted, that one holds in one's hands an obvious classic at the nexus of art and science. It's the feeling I had when I first read Primo Levi's *The Periodic Table*, or saw Irving Geis's protein structures, or looked at Peter S. Stevens's *Patterns in Nature*.

The meeting point here is of photography and surface science, the latter field itself a commons of chemistry, engineering and physics. Felice Frankel is an internationally recognized photographer, whose previous work has been in photographing landscape architecture (also surfaces, also natural and unnatural). George Whitesides is an outstanding chemist, versatile in the extreme, as savvy to the applied as he is to the pure. This modestly priced book contains about 60 photographs by Frankel, and short accompanying texts by Whitesides.

The photographs are startlingly beautiful. In fact, two readers of *On the Surface of Things* are likely to fall into this dialogue: "Did you see that photo of a ferrofluid, those weird spikes caught between surface tension and magnetism?" "Yes, I did. But

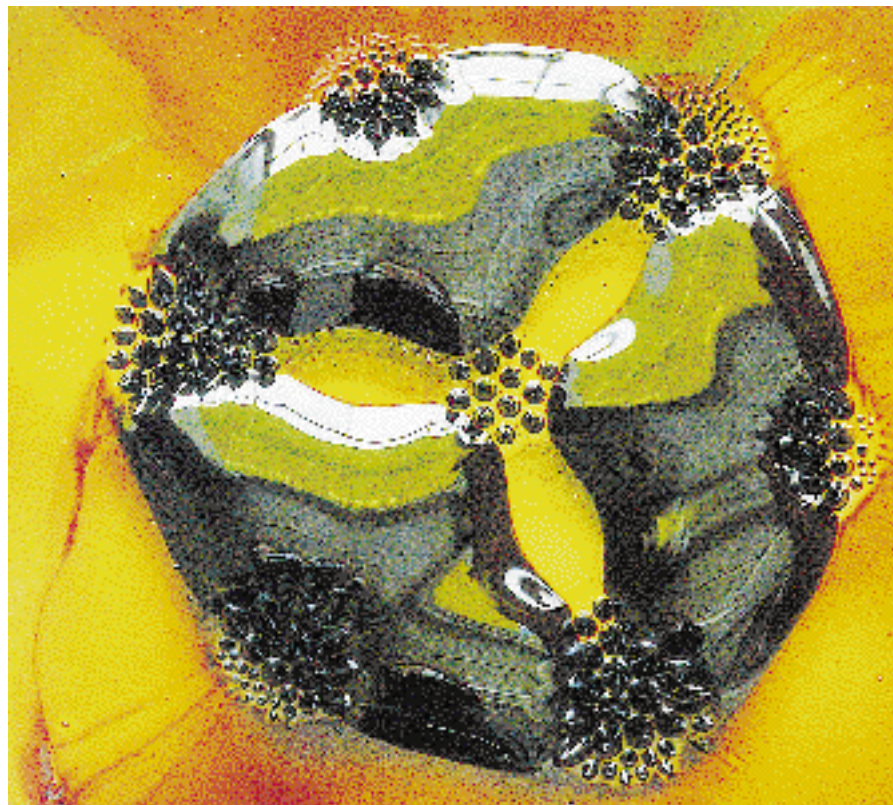
what I liked best were the mauve and red-brown rings in the Belousov–Zhabotinsky reaction!" "Really? Did you see that film chaotically peeling off a silicon crystal?"

At the end of the book are little paragraph notes by Frankel on the photographic techniques she used. These are written in an engaging personal style, and, once one discovers them, one is drawn to reading them in tandem with each image.

Frankel's spectacular photographs are the heart of this book, to be sure. But Whitesides also surprises. First, he succeeds (in too small a font, the only design failure of this book) in explaining complex matters of spectroscopy, electronics and the properties of materials without the impediment (to some) and the crutch (to others) of mathematics. Second, he writes of surfaces evocatively, really crafting prose poems around each image.

Art is no more an accident than science. The serious play of the mind, what Immanuel Kant called "the mutual relations of the imagination and the understanding" (and the hand, and the eye), is central to both human enterprises. It is then no surprise that a different way of representing (therefore seeing) an object of scientific investigation — through the artist's sensibility — should be of scientific value. This, just as much as the sheer beauty of the images, is the importance of what Frankel and Whitesides are doing. □

Ronald Hoffmann is in the Department of Chemistry, Cornell University, Ithaca, New York 14853-1301, USA.



Ferrofluid – "a gryphon in the world of materials: part liquid, part magnet".

Histone acetylation in chromatin structure and transcription

Michael Grunstein

The amino termini of histones extend from the nucleosomal core and are modified by acetyltransferases and deacetylases during the cell cycle. These acetylation patterns may direct histone assembly and help regulate the unfolding and activity of genes.

Nucleosomes, which fold chromosomal DNA, contain two molecules each of the core histones H2A, H2B, H3 and H4. Almost two turns of DNA are wrapped around this octameric core, which represses transcription *in vivo* and *in vitro*. But despite its similarity to a bead on a string, this building block of the chromosome is unexpectedly dynamic. The histone amino termini extend from the core, where they can be modified post-translationally by acetylation, phosphorylation and methylation, affecting their charge and function. Of such modifications, acetylation and deacetylation have generated most interest since gene activity was first correlated with histone acetylation¹. Further indications of this correlation emerged when an antibody recognizing ϵ -N-acetyl lysine was shown to bind chromatin enriched in a globin gene from chicken erythrocytes. Binding continued even when the gene was inactive², indicating that it is the potential for gene activity, rather than the transcriptional process itself, that is associated with hyperacetylated histones. As acetylation neutralizes the positively charged lysine residues of the histone N termini, decreasing their affinity for DNA, this might allow the termini to be displaced from the nucleosome, causing the nucleosomes to unfold and increasing access to transcription factors³⁻⁶.

How acetylation and deacetylation target specific genes or chromosomal domains has long been a mystery. Euchromatin and heterochromatin exhibit different acetylation patterns (euchromatin is the chromatin that decondenses during interphase of the cell cycle and which contains most of the genes coding for cellular proteins; heterochromatin is condensed even in interphase, often contains repetitive DNA, and is generally silent). Antibodies against specific acetylation sites in histone H4 have been used to show that potentially active euchromatin can be modified at all the H4 acetylable lysines (K5, K8, K12 and K16), whereas H4 in heterochromatin is hypoacetylated^{7,8}. None of the four sites seems to be acetylated on the inactive X chromosome in humans⁹, and *Drosophila* and yeast heterochromatin seems to be acetylated only on residue K12 (refs 10,11). But the resolution of these studies does not exclude the possibility that other patterns of acetylation are associated with subdomains of hetero- or euchromatin.

One example of an association between acetylation and a function other than gene activity is that of H4 diacetylation and assembly. Unlike histones in chromatin, newly synthesized H4 is preferentially acetylated at K5 and K12 (ref. 12), suggesting that this pattern is associated with assembly. But just as it has been unclear whether certain acetylation patterns are required to activate genes in euchromatin or to silence them in heterochromatin, we have not known whether histone H4 diacetylation is required for nucleosome assembly. This situation is rapidly changing. Not only have the consequences of altering acetylation sites in H3 and H4 been investigated, but many of the genes and enzymes responsible for histone acetylation and deacetylation have been identified. New tools have thus become available to investigate nucleosome assembly, the differences between heterochromatin and euchromatin, and the mechanisms of gene activation and repression.

Nucleosome assembly

In nucleosome assembly, binding of the H3/H4 tetramer to DNA is followed by assembly of H2A/H2B dimers. Newly synthesized histone H4 is diacetylated at K5/K12 in organisms as diverse as *Tetrahymena*, flies and humans. Newly synthesized H3 is also acetylated, albeit in a less conserved manner^{12,13}. How important is this acetylation in histone assembly? Chromatin assembly factor 1 (CAF1) is a complex of proteins p150, p60 and p48 from human cells which assembles newly synthesized H3 and H4 onto replicating DNA *in vitro*¹⁴. The complex is less capable of assembling histones isolated from nuclei, suggesting that the different acetylation pattern of chromosomal histones precludes assembly *in vitro*. CAF1 can bind H4 that is mono-, di- or triacetylated at K5, K8 and K12, or even unacetylated. Some H3 in the complex is monoacetylated, although about 60% is unacetylated¹⁵. Is K5/K12 diacetylation of H4 not essential for assembly? If K5/K12 diacetylation is important, why can these sites be deleted or mutated in yeast without strongly affecting cellular viability, given that defective nucleosome assembly can be lethal¹⁶⁻¹⁸?

These apparent inconsistencies partly reflect the redundancy of the H3 and H4 N termini in assembly of the H3/H4 tetramer. Either can be deleted without preventing nucleosome assembly in yeast cell extracts¹⁸ or even in *Xenopus* cell extracts¹⁹, but deletion of both in the same extract blocks assembly¹⁸. The H3 and H4 N termini can even be exchanged between the two proteins without strongly affecting assembly or viability. The effects of H4 N-terminal mutations on assembly must thus be examined in the absence of the H3 N-terminal tail. Even then, H4 K5, K8 and K12 must all be mutated simultaneously to prevent nucleosome assembly²⁰. The redundancy between these sites helps to explain CAF1's interactions with H3 and H4. CAF1 may recognize the H3/H4 histone complex as long as at least one H4 site is acetylated (Fig. 1), serving as a chemical tag or changing protein conformation to allow CAF1 to assemble the H3/H4 complex. Thus, H4 K5/K12 diacetylation may precede the pattern used for assembly by CAF1. Disruption of the CAF1 p160 counterpart in yeast does not block cell viability or nucleosome assembly^{24,25}, so other pathways of nucleosome assembly are likely to exist.

How H2A and H2B are assembled is less clear. A factor (NAP1) identified in humans, *Drosophila* and yeast^{21,22} facilitates the assembly of nucleosome arrays in an ATP-dependent manner²². This factor is bound to H2A and H2B in whole-cell extracts and moves from the nucleus in S phase (when much nucleosome assembly is taking place) to the cytoplasm in G2. Its action has not yet been shown to depend on DNA replication or histone acetylation, although some nucleosome assembly probably occurs *in vivo* even without DNA replication²³.

Genes that may be involved in the acetylation of histones for assembly have been identified. Cytoplasmic acetylation for histone assembly involves B-type histone acetyltransferases (HATs), as opposed to A-type HATs, which are nuclear and transcription related. The major B-type *HAT1* gene of yeast has been cloned and can acetylate residues 1-28 of H4 at K5 and K12 *in vitro*^{26,27},

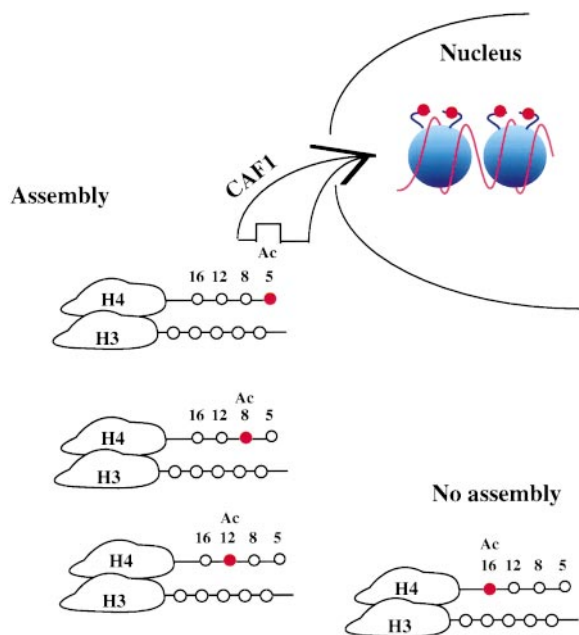


Figure 1 Recognition of the acetylated H3/H4 complex by human CAF1. A single acetyl group (Ac) on H4 K5, K8 or K12, but not K16, may allow CAF1 to recognize the complex. Multiply acetylated H4 K5/K8/K12 would also be recognized. The acetyl group may serve as a chemical tag or alter protein conformation to induce CAF1 binding. Unacetylated H3 would thus be incorporated into nucleosomes by way of H4 acetylation.

indicating that it is a catalytic subunit. HAT1 (M_r 42K) is associated with a second protein HAT2 (M_r 48K), which is required for full HAT1 activity and is similar in sequence to the p48 subunit of CAF1. The p48 subunit is also associated with the human deacetylase HD1 (ref. 28) and a similar histone deacetylase in flies²⁹. p48-type proteins may target catalytic subunits to histones, and HAT2 is indeed required for HAT1 to bind to an affinity resin containing the unacetylated H4 N terminus. As yeast extracts containing HAT1 acetylate only K12 in H4, negative regulators may restrict its specificity. Here too, disruption of HAT1 is neither lethal, nor appears to affect histone H4 acetylation *in vivo*. The redundancy of sites K5, K8 and K12 of histone H4 in nucleosome assembly may also extend to other B-type HATs which recognize these sites.

Heterochromatin

In *Drosophila* and yeast, heterochromatin is acetylated only at H4 K12, suggesting the importance of this modification in silencing^{10,11}. Human heterochromatin does not seem to be similarly acetylated⁹, however, and whereas mutations at H4 K16 completely disrupt silencing, similar alteration of K12 (or even simultaneous alteration of K5, K8 and K12) does not^{16,17,30,31}. Why is K12 acetylation not required for the integrity of heterochromatin? Either K12 and K16 acetylation may be redundant under certain conditions¹¹, or the explanation may lie in the condensed nature of heterochromatin. Yeast heterochromatin is packaged by specialized proteins that include the repressor/activator protein RAP1, the silencing information regulators (SIRs) 2, 3 and 4, and histones H3 and H4. These proteins are thought to interact as shown in Fig. 2, forming a structure that folds back onto subtelomeric regions to generate a condensed and protective chromosomal cap³². As HAT1 acetylation of K12 is the main B-type activity in yeast^{26,27}, all H4 may be preferentially acetylated at this site during chromatin assembly. Unlike euchromatin, assembled heterochromatin excludes many enzymes³³ and if acetyltransferases and deacetylases are among these, site 12 may remain preferentially acetylated (Fig. 2).

Mutations in the yeast histone deacetylases HDA1 and RPD3 increase general H3 and H4 acetylation levels *in vivo*³⁴ and would be

expected to unfold heterochromatin. Curiously, however, mutations in these enzymes increase repression of telomeric URA3 (refs 34–36). Similar hyper-repression of a gene adjacent to heterochromatin occurs when RPD3 in *Drosophila* is mutated³⁶. However, deacetylase disruption alters expression of many genes in yeast and mammals^{34,37,38}, and is likely to do so in *Drosophila*. In both yeast and flies, even small increases in certain components of heterochromatin can improve silencing and increase the extent of heterochromatin^{39–41}. Thus histone hypoacetylation in heterochromatin may still be a result of its condensed state.

Transcription

The core TATA promoter is recognized by the TATA-binding protein (TBP), TBP-associated factors (TAFs), and other proteins making up the proposed holoenzyme of RNA polymerase⁴². The activity of this basal transcription complex is enhanced by activator proteins, which recognize upstream activating sequences (UAS). In yeast, activator binding sites are usually nucleosome-free, and may be kept that way by specialized proteins (such as GRF2 at the GAL1 UAS^{43,44}), but TATA elements are often found in or near nucleosomes⁴³. How activators function, especially in the context of chromatin, has been a mystery.

Histone acetylation was known to affect transcription *in vivo*. Simulating acetylation by mutating lysines in H4 to uncharged residues, for example, can alter both nucleosome positioning^{45,46} and gene activity, often increasing uninduced transcription and reducing activated transcription^{46,47}. But actual acetylation of the histone N termini, in particular that of H4 (ref. 48), may trigger orderly nucleosomal disruption near TATA elements conducive to efficient transcription. How could this occur?

The HATs are likely to participate in gene activation. Histone acetyltransferase A in *Tetrahymena* is homologous to the yeast GCN5 protein which is required for activation of several genes in yeast and has intrinsic activity⁴⁹. GCN5 may function as part of a complex of proteins (including ADA2 and ADA3) connecting activators to the basal transcription machinery^{50–52}. In vertebrates, other transcription factors also have HAT activity. In particular, the homologous proteins CBP and p300 (often termed p300/CBP) link activators to the transcription machinery at promoters regulated by nuclear hormone receptors, the adenoviral oncoprotein E1A and other regulators⁵³. p300/CBP interacts with a second factor, pCAF, which, like p300/CBP, has intrinsic HAT activity^{54–56}. E1A represses many activators⁵⁷ and competes with pCAF for binding to p300/CBP, so pCAF and p300/CBP may associate to acetylate chromatin at certain promoters, activating transcription⁵⁴.

Recent findings have strengthened the link between histone acetylation and nucleosomal disruption in yeast. Tethering of TBP or a component of the RNA polymerase holoenzyme upstream of the TATA element by a sequence-specific DNA-binding protein was known to bypass the need for an activation domain, indicating that the activator may recruit the RNA polymerase holoenzyme to the downstream promoter⁵⁸. However, a protein associated with TBP (TAF130 in yeast, TAF230 in *Drosophila*, and TAF250 in humans) has intrinsic HAT activity⁵⁹. If TAF130 or GCN5 (or related proteins) dislodge histone N termini from nucleosomes at the promoter, tethering of these proteins by activators can explain how activators can disrupt nucleosomes even without a functional TATA element⁶⁰.

How histone deacetylases function at specific promoters has recently attracted considerable attention. An affinity matrix containing trapoxin, a deacetylase inhibitor, was used to isolate a human protein (HD1 or HDAC1) that co-immunoprecipitates histone deacetylase activity²⁸. HD1 is homologous to yeast RPD3, mutations in which alter expression of several genes^{34,37}. A second mammalian protein, HDAC2, is 85% identical to RPD3 and is associated with the transcriptional activator/repressor YY1 (ref. 61). A 350K histone deacetylase complex (HDA) from yeast nuclei

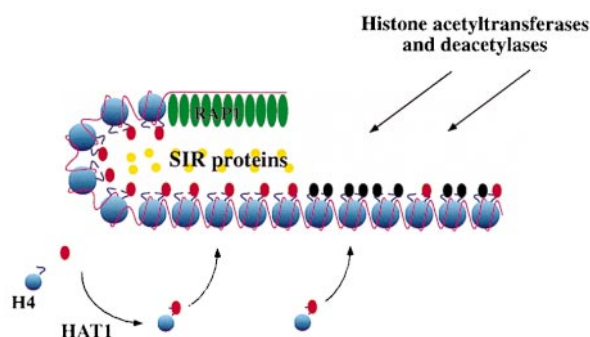


Figure 2 Preferential acetylation of H4 K12 in heterochromatin. Yeast heterochromatin is shown as a looped condensed structure formed by the interaction of RAP1, which binds to telomeric DNA, with SIR proteins, which interact with H3 and H4 (ref. 32). H4 K12 is thought to be acetylated by HAT1 during histone assembly so that the site is modified in both heterochromatin and euchromatin. Subsequent acetylation and deacetylation of H4 in euchromatin would generate H4 acetylated at residues K5, K8, K12 and K16 (red dots refer to acetylated K12; black dots to acetylated K5, K8 or K16), while heterochromatin blocks access to both histone acetyltransferases and deacetylases and so remains enriched in only K12.

containing three polypeptides, HDA1, HDA2 and HDA3, has also been purified to near homogeneity⁶². HDA1 is related to RPD3 and to three other histone deacetylases from yeast, HOS1, HOS2 and HOS3 (ref. 34). These enzymes probably act on histones *in vivo*, as their mutagenesis increases histone H3 and H4 acetylation³⁴. They may have overlapping but non-identical functions³⁴, targeting specific genes, chromosomal domains, or portions of the cell cycle.

How certain deacetylases act in a gene-specific manner is now becoming clear. Histone deacetylation should lead to repression, and a yeast histone H4, in which all four acetylation sites have been mutated to simulate hypoacetylation, represses every gene tested⁴⁷. Similarly, tethering HDAC2 to a GAL4 DNA-binding domain represses an adjacent gene even in the absence of the repressor YY1, suggesting that YY1 represses by tethering the histone deacetylase⁶¹. The yeast co-repressor SIN3, also known as RPD1, is thought to function in the same pathway as RPD3, because mutations in *caner* have similar yet non-additive effects on the same genes. These effects mimic those of histone acetylation-site mutations, increasing uninduced but decreasing induced activity^{37,63}. Even before mammalian RPD3 homologues were discovered, SIN3 homologues (mSIN3A and mSIN3B) were known to form a complex with the mammalian repressor Mad. Mad forms a complex with the protein Max to repress certain genes. Regulatory sequences would otherwise be activated by the Myc/Max heterocomplex^{64,65}, which recognizes the same E-box-related DNA sequence. Mad/Max and mSIN3 form a ternary complex, and tethering mSIN3 to DNA represses adjacent transcription, suggesting that Mad/Max tethers mSIN3 to the E box. The discovery of HD1 (HDAC1) in mammals indicated that this type of repression might occur by interaction between the histone deacetylase and mSIN3, deacetylating adjacent histones and allowing their tails to displace transcription factors⁶⁶. mSIN3A indeed interacts with the histone deacetylases HDAC1 and HDAC2 *in vivo* and this complex contains deacetylase activity. Similar Mad-mediated repression is prevented by the deacetylase inhibitors trapoxin or trichostatin A^{67,68}. Moreover, a fusion protein containing the mSIN3A domain that interacts with HDAC and the GAL4 DNA-binding domain represses a reporter gene. mSIN3A mutants lacking this domain can still repress, however, suggesting that interaction with the deacetylase is not the only pathway responsible for repression⁶⁸. The Mad repression complex is also associated with a large (270K) protein (N-CoR), which, together

with a related co-repressor (SMRT) is also involved in repression by the ligand-free receptors for thyroid hormone and retinoic acid. These receptors both heterodimerize with the 9-*cis*-retinoic-acid receptor, RXR, and require mSIN3 and histone deacetylase for repression. N-CoR and SMRT may interact directly with mSIN3⁶⁹⁻⁷¹.

Some aspects of this mechanism are highly conserved. The yeast protein UME6, which represses some genes involved in meiosis, may tether SIN3 and RPD3 to an upstream repressor DNA sequence. UME6 and SIN3 interact directly⁷², but there is little evidence that SIN3 interacts directly with histone deacetylases. Although HDAC1, mSIN3A and RbAp48 can be co-immunoprecipitated when expressed in insect cells, conserved insect proteins may hold the complex together⁶⁷. But other components of deacetylase complexes continue to emerge. Components of the RPD3-like human complex include HDAC1 and HDAC2, p48-like proteins (RbAp46 and RbAp48), which could target the deacetylase to histones, and other proteins (SAP18 and SAP30) of unknown function. SAP18 interacts directly with mSIN3 *in vitro*⁷³. Binding to specific target sequences may alter the components of some complexes, however, and only a small fraction of deacetylase complexes may contain RbAp48-like proteins. Figure 3 summarizes one mechanism of deacetylase mediated repression⁷⁴.

future directions

The conservation of histones, HATs and histone deacetylases indicates that modifications of nucleosome structure occur in all eukaryotes, although the mechanisms are still poorly understood. Whether HATs unfold histone N termini from nucleosomal DNA or displace nucleosomes at specific promoters, what histone modifications are required for efficient transcription, or how different acetyltransferases and deacetylases cooperate to produce them, are all still unclear. Moreover, acetylation experiments using free histones, rather than chromatin, *in vitro* can be misleading, and proteins other than histones may also be substrates for these enzymes *in vitro* or *in vivo*. High-mobility-group (HMG) proteins, for example, can be acetylated, and their deacetylation is inhibited by a histone deacetylase inhibitor⁷⁴.

Transcription factors that can activate or repress transcription may do so by associating with HATs or deacetylases as appropriate. Nuclear hormone receptors, for instance, with and without ligand, may bind to HATs (p300/CBP and pCAF) and to deacetylases (HDAC1), respectively.

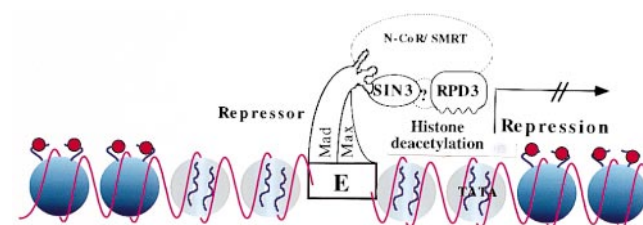


Figure 3 Transcription factors and repression by histone H3/H4 deacetylation. Repressors include Mad (when interacting with Max), UME6 or heterodimerized nuclear hormone receptors. The repressor recognizes its DNA-binding site and interacts directly with SIN3. The interaction of SIN3 with RPD3 (HDAC) may not be direct. Proteins such as SMRT and N-CoR form part of the complex at certain promoters, and other proteins such as SAP18, SAP30, RbAp46, RbAp48 may form part of the complex before or after it interacts with a DNA-bound repressor. Deacetylation would then allow the otherwise extended and acetylated histone N termini (red dots) to bind DNA, preventing access to transcription factors. It is assumed that repression can occur on either side of the repressor even when the repressor is not upstream of the transcription start site. E, E box.

Five histone deacetylases are known in yeast, suggesting that new ones will be found in higher eukaryotes. Are they involved in gene-specific repression, or do they act on larger chromosomal domains, perhaps at specific periods of the cell cycle? Which sites on which histones do they affect? And do they affect all promoters or chromosomal domains equally? Mutations in HATs, deacetylases and components of these complexes are now available in yeast, and a combination of genetics and biochemistry should clarify the relation between chromatin structure and gene activity. The results may have implications for human disease. The interaction of Mad and Max is important in tumour suppression⁷⁴, as is the integrity of the thyroid-hormone repressor. Moreover, drugs targeting histone deacetylases have been used against malaria and toxoplasmosis⁷⁵. This much is certain: histones are important regulators of gene activity. □

Michael Grunstein is in the Department of Biological Chemistry, UCLA School of Medicine and the Molecular Biology Institute, University of California, Los Angeles, California 90095, USA.

1. Allfrey, V. G. in *Chromatin and Chromosome Structure* (eds Li, H. J. & Eckhardt, R.) 167–191 (Academic, New York, 1977).
2. Hebbes, T. R., Thorne, A. W. & Crane-Robinson, C. A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7**, 1395–1402 (1988).
3. McGhee, J. D. & Felsenfeld, G. Nucleosome structure. *Annu. Rev. Biochem.* **49**, 1115–1156 (1980).
4. Gross, D. A. & Garrard, W. T. Nucleosome hypersensitive sites in chromatin. *Annu. Rev. Biochem.* **57**, 159–197 (1988).
5. Norton, V. G., Marvin, K. W., Yau, P. & Bradbury, E. M. Nucleosome linking number change controlled by acetylation of histones H3 and H4. *J. Biol. Chem.* **265**, 19848–19852 (1990).
6. Lee, D. Y., Hayes, J. J., Pruss, D. & Wolffe, A. P. A positive role for histone acetylation in transcription factor access to DNA. *Cell* **72**, 73–84 (1993).
7. Clarke, D. J., O'Neill, L. P. & Turner, B. M. Selective use of H4 acetylation sites in the yeast *Saccharomyces cerevisiae*. *Biochem. J.* **294**, 557–561 (1993).
8. O'Neill, L. P. & Turner, B. M. Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. *EMBO J.* **14**, 3946–3957 (1995).
9. Jeppesen, P. & Turner, B. M. The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell* **74**, 281–289 (1993).
10. Turner, B. M. *et al.* Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell* **69**, 375–384 (1992).
11. Braunstein, M. *et al.* Efficient transcriptional silencing in *Saccharomyces cerevisiae* requires heterochromatin histone acetylation pattern. *Mol. Cell Biol.* **16**, 4349–4356 (1996).
12. Sobel, R. E. *et al.* Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. *Proc. Natl Acad. Sci. USA* **92**, 1237–1241 (1995).
13. Kuo, M. *et al.* Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines. *Nature* **383**, 269–272 (1996).
14. Kaufman, P. D. *et al.* The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. *Cell* **81**, 1105–1114 (1995).
15. Verreault, A., Kaufman, P. D., Kobayashi, R. & Stillman, B. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell* **87**, 95–104 (1996).
16. Kayne, P. S. *et al.* Extremely conserved histone N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. *Cell* **55**, 27–39 (1988).
17. Megee, P. C., Morgan, B. A., Mittman, B. A. & Smith, M. M. Genetic analysis of histone H4: essential role of lysines subject to reversible acetylation. *Science* **247**, 841–845 (1990).
18. Ling, X. *et al.* Yeast histone H3 and H4 amino termini are important for nucleosome assembly *in vivo* and *in vitro*: redundant and position-independent functions in assembly but not in gene regulation. *Genes Dev.* **10**, 686–699 (1996).
19. Freeman, L., Kurumizaka, H. & Wolffe, A. P. Functional domains for assembly of histones H3 and H4 into the chromatin of *Xenopus* embryos. *Proc. Natl Acad. Sci. USA* **93**, 12780–12785 (1996).
20. Ma, X., Wu, J., Althelm, B., Schultz, M. C. & Grunstein, M. Histone H4 acetylation sites that are required for position *in vivo*. *Proc. Natl Acad. Sci. USA* (submitted).
21. Ishimi, Y. & Kikuchi, A. Identification and molecular cloning of yeast homolog of nucleosome assembly protein I which facilitates nucleosome assembly *in vitro*. *J. Biol. Chem.* **266**, 7025–7029 (1991).
22. Ito, T. *et al.* *Drosophila* NAP-1 is a core histone chaperone that functions in ATP-facilitated assembly of regularly spaced nucleosomal arrays. *Mol. Cell Biol.* **16**, 3112–3124 (1996).
23. Kim, U.-J., Han, M., Kayne, P. & Grunstein, M. Effects of histone H4 depletion on the cell cycle and transcription of *Saccharomyces cerevisiae*. *EMBO J.* **7**, 2211–2219 (1988).
24. Kaufman, P. D., Kobayashi, R. & Stillman, B. Ultraviolet radiation sensitivity and reduction of telomeric silencing in *S. cerevisiae* cells lacking chromatin assembly factor-I. *Genes Dev.* **11**, 345–357 (1997).
25. Enomoto, S. *et al.* RLF1, a subunit of yeast chromatin assembly factor-I, is required for telomeric chromatin function *in vivo*. *Genes Dev.* **11**, 358–370 (1997).
26. Kleff, S., Andrusis, E. D., Anderson, C. W. & Sternglanz, R. Identification of a gene encoding a yeast histone H4 acetyltransferase. *J. Biol. Chem.* **270**, 24674–24677 (1997).
27. Parthun, M. R., Widom, J. & Gottschling, D. E. The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. *Cell* **87**, 85–94 (1996).
28. Taunton, J., Hassig, C. A. & Schreiber, S. L. A mammalian histone deacetylase related to the yeast transcriptional regulator RPD3p. *Science* **272**, 408–411 (1996).
29. Tyler, J. K. *et al.* The p55 unit of *Drosophila* chromatin assembly factor I is homologous to a histone deacetylase-associated protein. *Mol. Cell Biol.* **16**, 6149–6159 (1996).
30. Johnson, L. M., Kayne, P. S., Kahn, E. S. & Grunstein, M. Genetic evidence for an interaction between SIR3 and histone H4 in the repression of the silent mating loci in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **87**, 6286–6290 (1990).
31. Thompson, J. S., Ling, X. & Grunstein, M. Histone H3 amino terminus is required for telomeric and silent mating locus repression in yeast. *Nature* **369**, 245–247 (1994).

32. Strahl-Bolsinger, S., Hecht, A., Luo, K. & Grunstein, M. SIR2 and SIR4 interactions differ in more and extended telomeric heterochromatin of yeast. *Genes Dev.* **11**, 83–93 (1997).
33. Thompson, J. S., Hecht, A. & Grunstein, M. Histones and the regulation of heterochromatin in yeast. *Cold Spring Harb. Symp. Quant. Biol.* **53**, 247–256 (1993).
34. Rundlett, S. E. *et al.* HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc. Natl Acad. Sci. USA* **93**, 14503–14508 (1996).
35. Vannier, D., Balderes, D. & Shore, D. Evidence that the transcriptional regulators SIN3 and RPD3, and a novel gene (SDS3) with similar functions, are involved in transcriptional silencing in *Saccharomyces cerevisiae*. *Genetics* **144**, 1343–1353 (1996).
36. De Rubertis, F. *et al.* The histone deacetylase RPD3 counteracts genomic silencing in *Drosophila* and yeast. *Nature* **384**, 589–591 (1996).
37. Vidal, M. & Gaber, R. F. RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **11**, 6317–6327 (1991).
38. Van Lint, C., Emiliani, S. & Verdin, E. The expression of a small fraction of cellular genes is changed in response to histone hyperacetylation. *Gene Exp. 5*, 245–253 (1996).
39. Locke, J., Kotarski, M. A. & Tartof, K. D. Dosage-dependent modifiers of position effect variegation in *Drosophila* and a mass action model that explains their effects. *Genetics* **120**, 181–198 (1988).
40. Renaud, H. *et al.* Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. *Genes Dev.* **7**, 1133–1145 (1993).
41. Hecht, A., Strahl-Bolsinger, S. & Grunstein, M. Spreading of transcriptional repression by SIR from telomeric heterochromatin. *Nature* **383**, 92–96 (1996).
42. Wilson, C. J. *et al.* RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. *Cell* **84**, 235–244 (1996).
43. Grunstein, M. Histone function in transcription. *Annu. Rev. Cell Biol.* **6**, 643–678 (1990).
44. Chasman, D. I. *et al.* A yeast protein that influences the chromatin structure of UASg and functions as a powerful auxiliary gene activator. *Genes Dev.* **4**, 503–514 (1990).
45. Roth, S. Y., Shimizu, M., Johnson, L., Grunstein, M. & Simpson, R. T. Stable nucleosome positioning and complete repression by the yeast $\alpha 2$ repressor are disrupted by amino-terminal mutations in histone H4. *Genes Dev.* **6**, 411–425 (1992).
46. Fisher-Adams, G. & Grunstein, M. Yeast histone H4 and H3 N termini have different effects on the chromatin structure of the GAL1 promoter. *EMBO J.* **14**, 1468–1477 (1995).
47. Durrin, L. K., Mann, R. K., Kayne, P. S. & Grunstein, M. Yeast histone H4 N terminal sequence is required for promoter activation *in vivo*. *Cell* **65**, 1023–1031 (1991).
48. Wan, J. S., Mann, R. K. & Grunstein, M. Yeast histone H4 and H4 N termini function through different GAL1 regulatory elements to repress and activate transcription. *Proc. Natl Acad. Sci. USA* **92**, 5664–5668 (1995).
49. Brownell, J. E. *et al.* Tetrahymena histone acetyltransferase A: a homologue to yeast Gcn5p linking histone acetylation to gene activation. *Cell* **84**, 843–851 (1996).
50. Horiuchi, J., Silverman, N., Marcus, G. A. & Guarente, L. ADA3, a putative transcriptional adaptor, consists of two separable domains and interacts with ADA3 and GCN5 in a trimeric complex. *Mol. Cell Biol.* **15**, 1203–1209 (1995).
51. Georgakopoulos, T., Gounalaki, N. & Thireos, G. Genetic evidence for the interaction of yeast transcriptional co-activator proteins GCN5 and ADA2. *Mol. Gen. Genet.* **246**, 723–728 (1995).
52. Candau, R., Zhou, J.-X., Allis, C. D. & Berger, S. L. Histone acetyltransferase activity and interaction with ADA2 are critical for GCN5 function *in vivo*. *EMBO J.* **16**, 555–565 (1997).
53. Chakravarti, D. *et al.* Role of CBP/p300 in nuclear receptor signalling. *Nature* **383**, 99–103 (1996).
54. Yang, X., Ogryzko, V. V., Nishikawa, J., Howard, B. & Nakatani, Y. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* **382**, 319–324 (1996).
55. Ogryzko, V. V., Shiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953–959 (1996).
56. Bannister, A. J. & Kouzarides, T. The CBP co-activator is a histone acetyltransferase. *Nature* **384**, 641–643 (1996).
57. Borrelli, E., Hen, R. & Chambon, P. Adenovirus-2 E1A products repress enhancer-induced stimulation of transcription. *Nature* **312**, 608–612 (1984).
58. Ptashne, M. & Gann, A. Transcriptional activation by recruitment. *Nature* **386**, 569–576 (1997).
59. Mizzen, C. A. *et al.* The TAF(II) 250 subunit of TFIID has histone acetyltransferase activity. *Cell* **87**, 1261–1270 (1996).
60. Gaudreau, L., Schmid, A., Blaschke, D., Ptashne, M. & Horz, W. RNA polymerase II holoenzyme recruitment is sufficient to remodel chromatin at the yeast PHO5 promoter. *Cell* **89**, 55–62 (1997).
61. Yang, W. M., Inouye, C., Zeng, Y., Bearss, D. & Seto, E. Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. *Proc. Natl Acad. Sci. USA* **93**, 12845–12850 (1996).
62. Carmen, A. A., Rundlett, S. E. & Grunstein, M. HDA1 and HDA3 are components of a yeast histone deacetylase (HDA) complex. *J. Biol. Chem.* **271**, 15837–15844 (1996).
63. Stillman, D. J., Dorland, S. & Yu, Y. Epistasis analysis of suppressor mutations that allow HO expression in the absence of the yeast SWI5 transcriptional activator. *Genetics* **136**, 781–788 (1994).
64. Ayer, D. E. *et al.* Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* **80**, 767–776 (1996).
65. Schreiber-Agus, N. *et al.* An amino-terminal domain of Mxi1 mediates anti-Myc oncogenic activity and interacts with a homolog of the yeast transcriptional repressor Sin3. *Cell* **80**, 777–786 (1995).
66. Wolffe, A. P. Histone-deacetylase: A regulator of transcription. *Science* **272**, 371–372 (1996).
67. Hassig, C. A., Fleischer, T. C., Billin, A. N., Schreiber, S. L. & Ayer, D. E. Histone deacetylase activity is required for full transcriptional repression by mSin3A. *Cell* **83**, 341–347 (1997).
68. Laherty, C. D. *et al.* Histone deacetylases associated with the mSin3 corepressor mediate Mad transcriptional repression. *Cell* **83**, 349–356 (1997).
69. Heinzel, T. *et al.* A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature* **387**, 43–48 (1997).
70. Alland, L. *et al.* Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. *Nature* **387**, 49–55 (1997).
71. Nagy, L. *et al.* Nuclear receptor repression mediated by a complex containing SMRT, mSin3A and histone deacetylase. *Cell* **83**, 373–380 (1997).
72. Kadosh, D. & Struhl, K. Repression by Ume6 involves recruitment of a complex containing Sin3 corepressor and Rpd4 histone deacetylase to target promoters. *Cell* **83**, 365–371 (1997).
73. Zhang, Y., Itratti, R., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. *Cell* **83**, 357–364 (1997).
74. Pazin, M. J. & Kadonaga, J. T. What's up and down with histone deacetylation and transcription? *Cell* **89**, 325–328 (1997).
75. Darkin-Rattray, S. J. *et al.* Apicidin: a novel antiparasitic agent that inhibits parasite histone deacetylase. *Proc. Natl Acad. Sci. USA* **93**, 13143–13147 (1996).

Acknowledgements. I thank A. Berk, J. Kadonaga, G. Felsenfeld, S. Rundlett and A. Carmen for comments.

Lunar accretion from an impact-generated disk

Shigeru Ida^{*†}, Robin M. Canup[†] & Glen R. Stewart[†]

^{*} Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Meguro-ku, Tokyo 152, Japan

[†] LASP, University of Colorado, Campus Box 392, Boulder, Colorado 80309-0392, USA

Although the mechanism by which the Moon was formed is currently unknown, several lines of evidence point to its accretion from a circumterrestrial disk of debris generated by a giant impact on the Earth. Theoretical simulations show that a single large moon can be produced from such a disk in less than a year, and establish a direct relationship between the size of the accreted moon and the initial configuration of the debris disk.

Many models have been proposed for formation of the Moon¹, but no one has succeeded in showing the formation satisfactorily. The popular “giant impact”^{2,3} model states that a Mars-sized protoplanet hit the proto-Earth and generated a circumterrestrial debris disk from which the Moon accreted. This model has been favoured as it may well account for the dynamical and geochemical characteristics of the Moon (large angular momentum of Earth–Moon system, depletion of volatiles and iron). Many hydrodynamic simulations (a smoothed particle method) have modelled the impact process^{4–7}. They calculated the impact between two large protoplanets with iron cores and silicate mantles and followed the orbital evolution of the debris after the impact for short timescales (on the order of a few orbital periods). It is found that an impact by a Mars-sized body usually results in formation of a circumterrestrial disk rather than direct formation of a clump. (This trend is most clear in recent simulations^{5–7}.) The disk mass is usually smaller than $2.5M_L$, where M_L is the present lunar mass ($0.0123M_\oplus$; $M_\oplus$ is the Earth mass). Most of the disk material is distributed near or interior to the radius a_R of the Roche limit ($\sim 2.9R_\oplus$, where $R_\oplus$ is the radius of the Earth) if the orbital angular momentum of the impact is $1\text{--}2J_{EM}$, where J_{EM} is the angular momentum of the present Earth/Moon system. Within and near the Roche limit, the tidal force of the Earth inhibits accretional growth.

In contrast, little has been done to simulate the accretional process of the Moon. The only published accretion calculation is that of Canup and Esposito⁸ with a gas dynamic approach. They approximated disk particles as particles in a box and tracked the evolution of the mass distribution function at individual regions of the disk, modelling velocity evolution, accretion and rebounding of

the disk particles. They showed that, in general, many small moonlets are formed initially rather than a single large moon and concluded that the simplest way to form the present-sized moon is to begin with at least a lunar mass of material outside the Roche limit. However, in gas-dynamic calculations it is difficult to include non-local effects such as radial migration of the disk material and global interaction between formed moons and the disk. The importance of the radial diffusion out from the Roche limit has been pointed out through analytical argument⁹.

Here we perform directly N -body simulations, which automatically include non-local effects, to investigate global lunar accretion processes. The sequence of accretion of the moon from an impact-generated disk might be as follows^{8,10} (see Fig. 1). Initially, the disk would probably be a hot, silicate-vapour atmosphere/torus^{6,7}. Solid particles condense owing to cooling of the disk, possibly after some radial migration¹⁰. Subsequent collisions and fragmentation of the particles would damp initially large orbital eccentricities and inclinations of the particles to moderate values in a few orbital periods. Our simulations start from this stage and follow the collisional evolution to a moon(s). On a longer timescale, one or more formed moons gradually migrate outwards by tidal interaction with the Earth^{8,10}, sweeping remnants. We do not pursue such long-term evolution here.

We present the results of 27 simulations with different initial disk conditions. We found that a single large moon, rather than multiple moons, is usually formed at similar distance from the proto-Earth in $100\text{--}1,000$ orbital periods (about a month to a year). We also found that the final moon mass is mostly determined by a simple function of initial total mass and angular momentum of the disk. To estimate

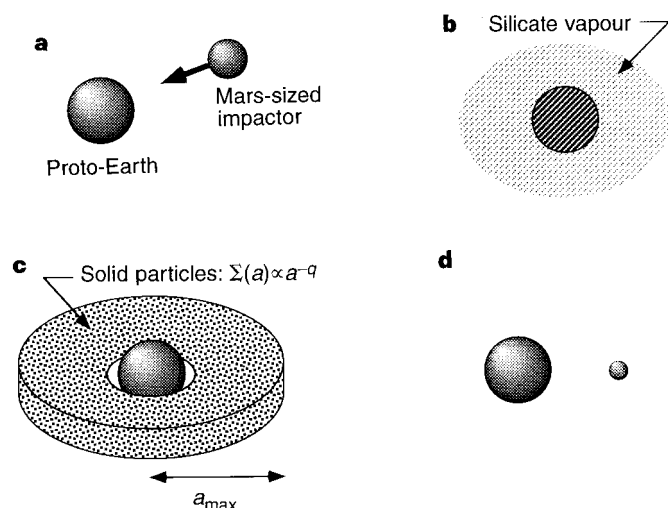


Figure 1 Schematic illustrations of the formation of the Moon by a giant impact: **a**, a Mars-sized body's impact on the proto-Earth; **b**, a hot, silicate vapour atmosphere/torus; **c**, a solid particle disk from which one or more moons accrete; **d**, outward migration of the formed moon(s) by tidal interaction with the Earth. We adopted stage **c** as the initial conditions for N -body simulations. The particle disk is modelled as follows. The disk consists of solid particles with a power-law size distribution as $n(m)dm \propto m^{-p}dm$, where m is mass of the particles. The surface density of the disk is given by $\Sigma(a) \propto a^{-q}$ for $0.35a_R < a < a_{\max}$, where a is the semimajor axis. The orbits of the disk particles are integrated by a fourth-order hermitian integrator¹⁶ with a hierarchical individual time step¹⁷, calculating all gravitational interactions between the particles, in geocentric cartesian coordinates. We take out particles from the system if they collide with the Earth or are scattered into hyperbolic orbits. We adopt the accretion criteria of Canup and Esposito¹¹ (see text).

the final moon mass, we do not need to know the details of initial mass, size and velocity distributions of the disk particles. The predicted moon mass from the disks obtained by the previous impact simulations might be as large as the present lunar mass in some cases. However, we cannot make a definitive conclusion at present, as the previous impact simulations did not provide enough data about the disk angular momenta. Improved simulations are needed to provide total mass and angular momentum of the disk. The combination of more refined N -body and impact simulations would clarify whether a giant impact could indeed have produced the Moon or not.

Model description

We simulated the formation of a moon from a (three-dimensional) circumterrestrial debris disk initially consisting of 1,000–2,700 particles with mass $m \approx 10^{-5}$ to $10^{-2} M_L$ (see Fig. 1), assuming that solid particles had condensed and attained such sizes through accretion. (In the inner part of the disk, the particles might remain very small, as accretion becomes increasingly inhibited inside the Roche limit. Furthermore, the disk material might remain liquid owing to the longer cooling time in the inner part. We will comment on these effects later.)

We calculated disks with as many different initial conditions as possible, as we do not have enough knowledge about disk conditions after the vapour/liquid phase and initial collisional evolution. The parameters we examined are summarized in Table 1, where we show 19 runs of the 27 simulations for which we retained detailed output data. As shown below, the final outcome of accretion has only a weak dependence on the details of conditions of a starting disk. We scale the orbital radii by the Roche radius defined by $a_R = 2.456(\rho_{\oplus}/\rho)^{1/3} R_{\oplus}$ where $(\rho_{\oplus}/\rho)$ is the ratio of the internal density of the Earth to that of the disk particles. For disk

particles with $\rho = 3.34 \text{ g cm}^{-3}$ (the bulk density of the Moon), a_R is located at about $2.9R_{\oplus}$. Using a_R , the physical radii of disk particles with mass m are given by $R = (1/2.456)(m/M_{\oplus})^{1/3} a_R$, independent of $(\rho_{\oplus}/\rho)$.

Near the Roche radius, tidal forces of the proto-Earth affect whether colliding particles rebound or accrete. Within $\sim 0.8a_R$, tidal forces preclude accretion, whereas in the transitional zone, $0.8\text{--}1.35a_R$, limited accretional growth can occur¹¹. Exterior to this zone, accretion is largely unaffected by tidal forces. This transitional zone will be referred to as the Roche zone. We adopt here the accretional criteria of Canup and Esposito¹¹, which include this transition in addition to the impact velocity condition that, for accretion, the calculated rebound velocity must be smaller than some critical value corresponding to the (mutual) surface escape velocity¹¹. If the colliding bodies in our simulation satisfy the criteria, we produce a merged body, conserving momentum. If not, the bodies rebound with given restitution coefficients (Table 1).

Characteristics of moon accretion

Below we present the results from several of the 27 disk simulations that were calculated. In most of the simulations, a single large body is formed near the Roche radius. In Figs 2 and 3, we show snapshots of the results for the disks with initial mass $M_{\text{disk}} = 0.03M_{\oplus}$ ($= 2.44M_L$). The unit of time is the keplerian rotation time at a_R , which is $\sim 7 \text{ h}$; $t = 100$ realistically corresponds to 1 month. Figure 2 shows a centrally confined disk case (run 4 in Table 1) in which the semimajor axes of all the particles are initially within the Roche radius, whereas Fig. 3 is a rather extended disk case (run 9). The extension of a disk is indicated by $J_{\text{disk}}/M_{\text{disk}}$, where J_{disk} is the total angular momentum of the starting disk. For the disks in Figs 2 and 3, $J_{\text{disk}}/M_{\text{disk}}$ are $0.692 \sqrt{GM_{\oplus}a_R}$ and $0.813 \sqrt{GM_{\oplus}a_R}$, respectively.

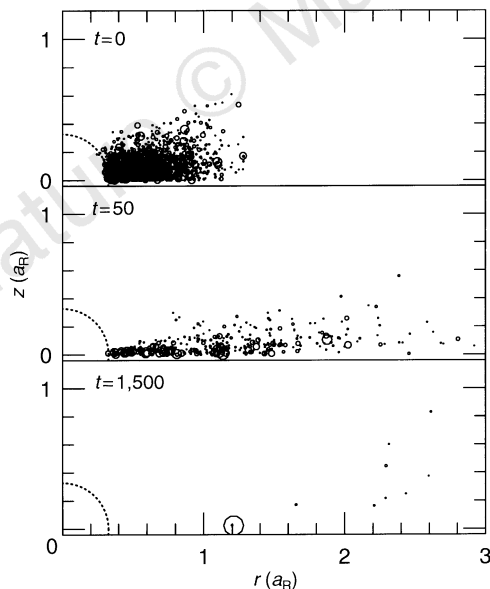


Figure 2 Snapshots of disk particles plotted in geocentric cylindrical coordinates (r, z). (Particles at negative z are plotted at $|z|$.) The units of length and time are the Roche limit radius, a_R , and Kepler time, T_{Kep} at a_R ($\sim 7 \text{ h}$). The solid and dotted circles are disk particles and the Earth, respectively. The sizes of the circles indicate physical sizes. The snapshots here are the result of run 4 in Table 1. The mean specific angular momentum, $J_{\text{disk}}/M_{\text{disk}}$, is initially $0.692 \sqrt{GM_{\oplus}a_R}$. At $t = 1,500$ the moon has mass $0.40M_L$, semimajor axis $1.20a_R$, eccentricity 0.09 and inclination (radian) 0.02. The second body's mass is only $0.025M_L$. The masses ejected from the system (M_{∞}) and that hit the Earth are $0.026M_L$ and $1.95M_L$, respectively.

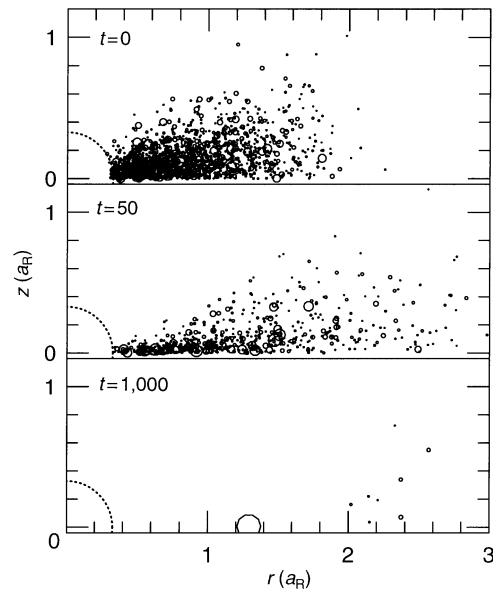


Figure 3 The same snapshots as in Fig. 2 but for run 9 of a more extended disk ($J_{\text{disk}}/M_{\text{disk}} = 0.813 \sqrt{GM_{\oplus}a_R}$). At $t = 1,000$ the largest moon mass is $0.71M_L$.

If all the disk material is at a_R , $J_{\text{disk}}/M_{\text{disk}} = \sqrt{GM_{\oplus}/a_R}$. The most recent impact simulations⁷ suggest $0.6\sqrt{GM_{\oplus}/a_R} \leq J_{\text{disk}} \leq 0.9\sqrt{GM_{\oplus}/a_R}$ (see the last section).

The middle panels of each figure show the first stage of disk evolution where the disk shrinks in the z -direction owing to collisional damping and the disk diffuses radially by angular momentum transfer. Density waves develop in the inner region with high surface density when the disk scale height becomes so small that the disk self-gravity becomes important. The waves accelerate the angular momentum transfer and the spreading of the disk material, as expected⁹. The waves are transient (duration is only ~ 10 rotational periods) because disk materials are quickly spread out so that the disk becomes gravitationally stable again. Furthermore, as shown below, large bodies immediately form near the Roche zone and soon start to perturb the entire disk. Hence, (smooth) disk evolution as an accretion disk terminates before the disk converges to some common geometry. We found that the dominant wave patterns look like several spiral arms. The relatively large number of arms might be caused by the relatively large ratio of inner radius to outer radius of the disk (annulus) that is required by the finite radius of the Earth at the centre of the disk. However, the spiral patterns are not always clear as the number of particles in our present simulations is insufficient to resolve them accurately. As the self-gravitational instability is regulated by surface density but not by each particle size¹², the density waves would develop similarly even if we were to start with more realistic, smaller particles, whereas transfer of angular momentum associated with collisions would diminish. As the density waves have an important role in mass transfer from the Roche limit, higher-resolution N -body simulations are required for future studies.

Near or outside the Roche zone accretion starts, whereas particles remain small in the inner region because accretion is tidally inhibited inside the Roche zone (see the middle panels of Figs 2 and 3). Note that substantial mass is transferred from inside the Roche zone to start accretion, in particular, in Fig. 2. As the orbital frequency, surface density, and hence collision rate decrease with

distance from the Earth, a large body is first formed near the Roche zone (innermost region of the accretion-permitted region). The large body has relatively small velocity dispersions (eccentricity and inclination), which is due to energy equipartition through collisions and gravitational scattering¹³. This facilitates rapid growth of the large body¹³. The largest bodies have attained 90% of the final masses already at $t = 150$ and 250 in the runs in Figs 2 and 3, respectively.

The increasing gravity of the largest body eventually perturbs almost the entire disk. This can be understood from the following argument: by the balance between gravitational/collisional stirring and dissipation due to inelastic collisions, the velocity dispersion, v , of the disk particles grows to about the surface escape velocity v_{esc} of the dominant body¹⁴, which is given by $(2Gm/R)^{1/2} \approx 2(m/M_{\oplus})^{1/3}(r/a_R)^{1/2}v_{\text{Kep}}$, where r is distance from the Earth and v_{Kep} is the Kepler velocity at r . The particles in the width $\Delta r \approx 2er \approx 2vr/v_{\text{Kep}}$ cross the large moon's orbit. Substituting $m \approx M_L$ and $r \approx a_R$, we find $\Delta r \approx a_R$. Hence, unless the disk is substantially extended or the final moon mass is much smaller than M_L , the entire disk is perturbed by the largest body's gravity, so that the other disk particles are 'cleaned up' (scattered onto the Earth or into hyperbolic orbits or accreted by the largest body). Thereby a single moon on a nearly circular orbit remains near the Roche zone as is shown in the bottom panels of Figs 2 and 3.

The other calculations with different initial conditions show a quite similar outcome, as long as an initial disk is not very extended: a single large moon on a nearly circular orbit. The final position of the large moon is $0.9-1.6a_R$ ($2.6-4.6R_{\oplus}$). The variety of the final position is caused by recoil of the 'clean-up'. If tidal interaction with the Earth were included, the formed moon would gradually migrate outward on a much longer timescale, presumably sweeping up any remaining smaller outer debris^{8,10}. (Inner remaining bodies within $0.7-0.8a_R$ (co-rotation radius) would gradually migrate to the Earth⁸.)

Figure 4 shows the evolution of a very extended disk ($J_{\text{disk}}/M_{\text{disk}} = 0.985\sqrt{GM_{\oplus}/a_R}$; run 13) where the disk width is

Table 1 Input parameters and results of the simulations

$M_{\text{disk}} = 2.44; p = 1.5$													
Run	$J_{\text{disk}}/M_{\text{disk}}$	J_{disk}	a_{max}	q	N	ϵ_n	a	M	M'	M_{∞}	J	J'	J_{∞}
1	0.670	0.300	0.95	5	1,500	0.01	1.08	0.193*	0.356	0.010	0.036	0.084	0.001
2	0.670	0.300	0.95	5	1,500	0.5	1.69	0.216*	0.248	0.065	0.051	0.060	0.017
3	0.690	0.308	0.95	4	1,000	0.01	0.87	0.252	0.437	0.016	0.043	0.094	0.007
4	0.692	0.309	0.95	4	2,000	0.01	1.20	0.400	0.464	0.026	0.080	0.099	0.005
5	0.722	0.323	1.25	3	1,000	0.01	0.88	0.437	0.484	0.109	0.074	0.087	0.044
6	0.722	0.323	1.25	3	1,000	0.5	1.32	0.281	0.391	0.153	0.059	0.090	0.037
7	0.767	0.343	1.25	3	1,500	0.01	1.55	0.315	0.411	0.0005	0.070	0.100	0.0001
8	0.794	0.355	1.25	3	2,700	0.01	1.21	0.751	0.776	0.006	0.151	0.159	0.002
9	0.813	0.363	1.50	2	1,500	0.01	1.24	0.712	0.744	0.016	0.144	0.154	0.031
10	0.823	0.363	1.50	2	1,000	0.5	0.79	0.260*	0.636	0.178	0.042	0.151	0.046
11	0.834	0.373	1.80	2	1,000	0.01	1.24	0.666	0.721	0.212	0.135	0.151	0.079
12	0.891	0.398	2.00	2	1,000	0.01	1.87	0.572*	0.609	0.140	0.142	0.155	0.041
13	0.958	0.428	2.00	1	1,000	0.01	1.98	0.631*	0.646	0.198	0.162	0.226	0.040
14	0.977	0.437	2.00	1	1,000	0.01	1.12	1.046	1.122	0.299	0.203	0.225	0.099
$a_{\text{max}} = 1.25; q = 3$													
Run	$J_{\text{disk}}/M_{\text{disk}}$	J_{disk}	M_{disk}	p	N	ϵ_n	a	M	M'	M_{∞}	J	J'	J_{∞}
15	0.738	0.323	2.44	0.5	1,000	0.01	0.93	0.505	0.685	0.120	0.089	0.141	0.040
16	0.757	0.451	3.24	1.5	1,000	0.5	1.19	0.536	0.631	0.287	0.107	0.132	0.075
17	0.767	0.457	3.24	1.5	1,500	0.01	1.45	0.657	0.694	0.088	0.142	0.153	0.031
18	0.768	0.344	2.44	1.0	1,000	0.01	1.53	0.515	0.668	0.008	0.117	0.153	0.003
19	0.778	0.232	1.62	1.5	1,000	0.01	0.91	0.251*	0.444	0.001	0.044	0.096	0.004

Input parameters and output data of 19 runs of the 27 simulations for which we retained detailed output data. Snapshots of runs 4, 9 and 13 are shown in Figs 2, 3 and 4, respectively. Input parameters are $J_{\text{disk}}, M_{\text{disk}}, q, p, N$ and ϵ_n , which are angular momentum, mass and outer edge of an initial disk, power indices of surface density distribution ($\Sigma(a) \propto a^{-q}$, where a is the semimajor axis of the disk particles) and size distribution ($n(m) \propto m^{-p}$, where m is the mass of the particles), total number of disk particles and restitution coefficient in the normal direction, respectively. Units of angular momentum, mass and distance are the present angular momentum of the Earth/Moon system ($J_{\text{EM}} = 3.5 \times 10^{41} \text{ g cm}^2 \text{ s}^{-1}$), the present lunar mass M_L and the Roche radius ($a_R = 2.9R_{\oplus}$), respectively. As $M_L = 0.0123M_{\oplus}$, $M_{\text{disk}} = 1.62, 2.44$ and 3.24 correspond to $0.02M_{\oplus}, 0.03M_{\oplus}$ and $0.04M_{\oplus}$, respectively. The second column, $J_{\text{disk}}/M_{\text{disk}}$, is the specific angular momentum of the disk in units of $\sqrt{GM_{\oplus}/a_R}$ (not J_{EM}/M_L). The normal restitution coefficient, ϵ_n , is set to be a constant, 0.01 or 0.5, except for collisions with very small velocity, where ϵ_n approaches unity. The tangential restitution coefficient, ϵ_t , is fixed at unity. (The non-slip case with inclusion of particle spins is a subject for future research.) Note that for even $\epsilon_n = 0.01$, collisions are not very dissipative. In most cases, initial eccentricities and inclinations (radian) of the disk particles are given by a Rayleigh distribution with a mean value of 0.3. (Different initial mean eccentricity and inclination do not change the results.) Columns 8 to 14 are output data. M and J are the mass and orbital angular momentum of the largest moon (a is its semimajor axis). M' and J' are the same quantities but including all bound debris outside the largest moon's orbit, which might finally accrete to the largest moon through tidal orbital migration. The same quantities for ejected particles are M_{∞} and J_{∞} . In column 9 (M), cases where the second body is larger than 30% of the largest body are indicated by an asterisk.

larger than $2a_R$. As expected from the above discussion ($\Delta r \approx a_R$), two large moons are formed in this case. We found that for $J_{\text{disk}}/M_{\text{disk}} \geq 0.85 \sqrt{GM_{\oplus} a_R}$, multiple moons are more common (see Fig. 5). The variety in final positions of the large moon is larger ($0.8\text{--}2.0a_R$) owing to interaction between multiple large moons. In these very extended disk cases it would be important to study the coalescence mechanism of the large moons during the subsequent tidal orbital migration, as discussed by Canup and Esposito⁸.

The accretion timescale and final mass of the moon depend on initial conditions. For a large restitution coefficient and a centrally confined disk, the accretion time is long because accretion is not favoured by the collisions with a large coefficient or those in inner regions. The final mass is in general larger in more extended disks as suggested by Figs 2 and 3 ($0.50M_L$ in Fig. 2 but $0.71M_L$ in Fig. 3). We next consider what determines the final mass.

Before that, we comment on effects of liquidity of the disk particles. In inner disk regions a hot environment might be maintained^{16,7,10}, so that disk materials might be liquid (or possibly vapour). Furthermore, energy release during collisions might melt the disk particles. During collisions the kinetic energy of relative motion is translated into heat energy. The specific dissipative energy would be $\sim e^2 v_{\text{kep}}^2 \approx 200(e/1.0)^2 (r/a_R) \text{ (J g}^{-1}\text{)}$, which corresponds to a temperature increase of $\sim 200(e/0.1)^2 (r/a_R) \text{ K}$. As small particles tend to have large eccentricities (>0.3) as a result of the energy equipartition, they can melt totally during one collision (melting temperature of silicate is 1,200–1,300 K). Such melting might lead to tidal disruption of the particles inside the Roche zone to form a fluid disk. Then the fluid ‘blobs’ might be in a thick vapour or in a thick swarm of other ‘blobs’ so that drag from the vapour or very inelastic collisions with the other ‘blobs’ would prevent their eccentricities from getting too high, which might affect the ‘clean-up’ process. However, the tidal torque exerted by the moon would push the inner disk to the Earth even if the eccentricities of the ‘blobs’ were damped. The moon has strong resonant interactions (2:1, 3:2, 4:3, ...) with the disk in the region of $\geq 0.6a$, where a is the semimajor axis of the Moon, and the disk in the region can easily be cleared, even if the disk is composed of gas¹⁵. This suggests that the tidal torque of the moon at $\sim a_R$ might be able to push the disk material on to the Earth surface ($0.34a_R$), or at least to well inside the co-rotation radius at $0.7\text{--}0.8a_R$ (the material inside the co-rotation radius would migrate inward by tidal interaction with the Earth). Thus, the melting of the disk particles would not affect the ‘clean-up’ process we showed, although the timescale to accomplish it might become longer.

Mass of the moon

On the basis of the result obtained by N -body simulations, we can derive a simple formula to obtain the largest moon’s mass. It is a function of only two parameters: initial disk mass, M_{disk} , and angular momentum, J_{disk} . Conservation of angular momentum gives

$$J_{\text{disk}} \approx M \sqrt{GM_{\oplus}(1-e_1^2)a_1} + (M_{\text{disk}} - M - M_{\infty}) \sqrt{GM_{\oplus}(1-e_2^2)a_2} + M_{\infty} \sqrt{GM_{\oplus}(1-e_3^2)a_3} \quad (1)$$

where M , $(M_{\text{disk}} - M - M_{\infty})$ and M_{∞} are the masses of the largest moon, mass accreted to the Earth and ejected mass, respectively, and e_k , a_k and $\sqrt{GM_{\oplus}(1-e_k^2)a_k}$ ($k = 1, 2, 3$) are their orbital eccentricities, semimajor axes and specific angular momenta. Here we assumed that the total mass of remnant small bodies is negligible. The perigee distance $(1-e_3)a_3$ is $\sim a_1$, as the escaping bodies are usually scattered by the largest moon. In contrast, $(1-e_2)a_2 \leq R_{\oplus} (\sim 0.34a_R)$; we take $(1-e_2)a_2 \approx 0.3a_R$. From our simulations we find typical values of: $e_1 < 0.1$, $a_1 \approx 1.3a_R$, $e_2 \approx 0.2\text{--}0.3$ and $e_3 \approx 1$. Substituting these into

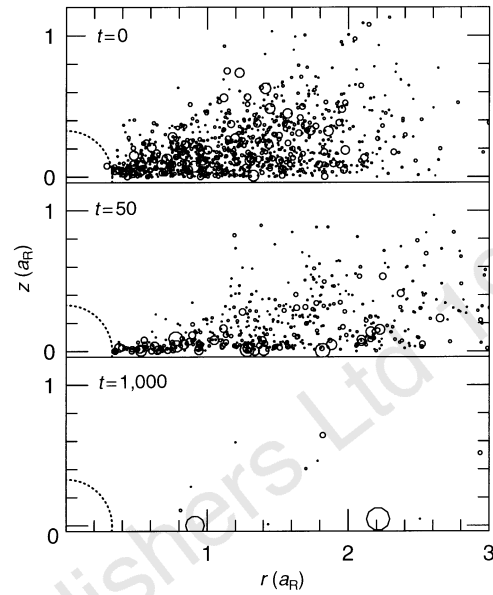


Figure 4 The same snapshots as in Fig. 2 but for run 13 of a very extended disk ($U_{\text{disk}}/M_{\text{disk}} = 0.958 \sqrt{GM_{\oplus} a_R}$). In this case two large moons are formed. At $t = 1,000$ the largest moon has mass $0.63M_L$ and semimajor axis $1.98a_R$, whereas the second has $0.39M_L$ and $0.93a_R$.

equation (1), we obtain

$$M \approx \frac{1.9J_{\text{disk}}}{\sqrt{GM_{\oplus} a_R}} - 1.15M_{\text{disk}} - 1.9M_{\infty} \quad (2)$$

Ejected mass M_{∞} is negligible in confined disks with small $J_{\text{disk}}/M_{\text{disk}}$ ($M_{\infty} \leq 0.05M_{\text{disk}}$ for $J_{\text{disk}}/M_{\text{disk}} \leq 0.8 \sqrt{GM_{\oplus} a_R}$), whereas it often takes a large value for more extended disks. The simulations with larger $J_{\text{disk}}/M_{\text{disk}}$ eject more material out of Earth orbit because disk material in shallow potential of the Earth (in the outer part) can easily be ejected and because gravitational interactions with multiple moons are more effective at pumping up the energy of the smaller particles. The ejected mass also tends to take a large value when the restitution coefficient is large.

In Fig. 5 we compare equation (2) (with both sides divided by M_{disk}) with N -body simulation results. The final moon mass in the N -body simulations is the value at which the residual disk mass becomes sufficiently small compared with the moon mass or disk evolution becomes very slow (usually at $t = 500\text{--}2,000$ rotational periods). The simulation results include large varieties of initial disk conditions, final moon masses and restitution coefficient. The lines of equation (2) with $M_{\infty} = 0$ and $0.05M_{\text{disk}}$ are drawn. For $J_{\text{disk}}/M_{\text{disk}} \leq 0.8 \sqrt{GM_{\oplus} a_R}$, the simulation results are consistent with the theoretical lines with small M_{∞} . The theoretical estimate would exhibit a smaller slope and would become more consistent with the simulation results if we allowed M_{∞} to increase systematically with $J_{\text{disk}}/M_{\text{disk}}$ as mentioned above. Larger M_{∞} inhibits the formation of a large moon, as much more mass than M_{∞} falls on to the Earth to compensate the angular momentum loss.

For disks with $J_{\text{disk}}/M_{\text{disk}} \leq 0.8 \sqrt{GM_{\oplus} a_R}$, which would in practice be important, the third term in the right-hand side of equation (2) is usually small compared with the other terms (if we adopt a non-slip rebounding condition, $e_i < 1$, M_{∞} might become smaller). Therefore, we conclude that the final moon mass is determined mostly by two parameters, J_{disk} and M_{disk} , but not by the other detailed initial conditions or rebounding conditions (their effects are included in a small correction term, M_{∞}). This conclusion allows

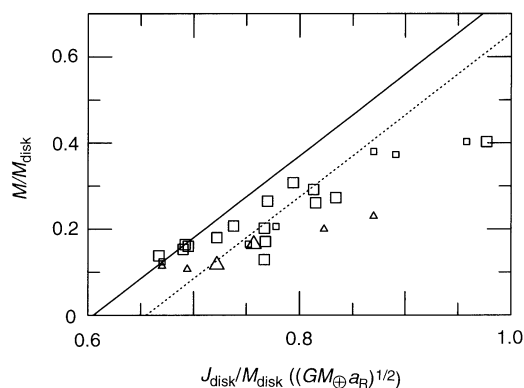


Figure 5 Final moon mass, M , is plotted on the $M/M_{\text{disk}} - J_{\text{disk}}/M_{\text{disk}}$ plane. Smaller $J_{\text{disk}}/M_{\text{disk}}$ cases are confined disks with smaller a_{max} and larger q . The solid line is the analytical estimate by equation (2) with $M_{\infty} = 0$; the dotted line is that with $M_{\infty} = 0.05M_{\text{disk}}$. Squares are N -body results with $\epsilon_n = 0.01$ and triangles are N -body results with $\epsilon_n = 0.5$. The small squares and triangles are the cases where the mass of a second moon is larger than 30% of the largest moon's mass. In these cases we plotted the sum of the largest and second moons' masses. Here we plotted the results of all 27 simulations.

us to predict easily the outcome of the complicated accretion process in the protolunar disk.

Application to the results of previous impact simulations

From the above results we estimate the expected moon masses that would result from the impact-generated disks obtained by previous hydrodynamic simulations. Unfortunately, the previous impact simulations^{4–7} did not present J_{disk} , although M_{disk} was presented in many cases. Here we roughly estimated J_{disk} from the published data^{4,5}: surface density histograms⁴ and a table of the initial disk masses within and exterior to the Roche limit⁵. The estimated $J_{\text{disk}}/M_{\text{disk}}$ is distributed from $0.7\sqrt{GM_{\oplus}a_R}$ to $1.1\sqrt{GM_{\oplus}a_R}$, although our estimates include a large uncertainty. Figure 4 suggests that $M/M_{\text{disk}} = 0.15–0.4$ for such $J_{\text{disk}}/M_{\text{disk}}$. Because the disks produced by the previous impact simulations have $M \leq 2.5M_L$, it seems possible for some disks to yield a lunar-sized moon. However, results of recent higher-resolution simulations⁷ ($1M_L$ is represented by ~ 300 smoothed particles) suggest that the resolution of the previous works ($1M_L \approx 37$ particles) might not have been adequate to describe the initial disk mass distribution. The higher-resolution disks generated tend to be more centrally confined and, equivalently, to have smaller $J_{\text{disk}}/M_{\text{disk}}$ (presumably, $J_{\text{disk}}/M_{\text{disk}} \approx 0.6–0.9\sqrt{GM_{\oplus}a_R}$ for impacts of $\leq 2J_{\text{EM}}$), although the higher-resolution simulations⁷ did not provide specific data about J_{disk} and M_{disk} . In this case we would predict a smaller final moon mass for similar M_{disk} .

In general, impacts with higher angular momentum produce more massive and extended disks^{4–6} that provide more material outside the Roche zone. We can predict the formation of a moon

with at least a lunar mass from most of the disks produced by the impacts with $\geq 2J_{\text{EM}}$, for relatively small M_{∞} . However, no reasonable means to rid the Earth/Moon system of this excess angular momentum has yet been proposed. This problem was pointed out by Canup and Esposito⁸. Furthermore, the present N -body simulations show that accretion from an extended disk often results in large M_{∞} , which inhibits the formation of a large moon as mentioned above and usually produces multiple moons, where we are faced with another problem on the coalescence of the multiple moons⁸.

In this work we have identified a direct relation between the initial mean specific angular momentum of a disk and the fraction of the disk mass that eventually becomes incorporated into the single large resulting moon. This result gives an important clue to the problem of the Moon formation (although more detailed N -body simulations might be needed to examine wider disk parameter ranges with more realistic, smaller starting particles). Improved impact simulations are needed to provide more detailed initial disk conditions, in particular J_{disk} and M_{disk} . Furthermore, angular momentum and mass transport might be important during the evolution from the initial vapour state to a particulate disk¹⁰. This problem should be addressed by some new fluid-dynamical simulations (or fluid/particles hybrid simulations). The combination of these simulations with different approaches is needed to clarify the formation of the Moon. $\square$

Received 21 March; accepted 1 August 1997.

1. Boss, A. P. & Peale, S. J. in *Origin of the Moon* (eds Hartmann, W. K., Phillips, R. J. & Taylor, G. J.) 59–102 (Lunar and Planetary Inst., Houston, 1984).
2. Hartmann, W. K. & Davis, D. R. Satellite-sized planetesimals and lunar origin. *Icarus* **24**, 504–515 (1975).
3. Cameron, A. G. W. & Ward, W. R. The origin of the Moon. *Proc. Lunar Planet. Sci. Conf.* **7**, 120–122 (1976).
4. Benz, W., Slattery, W. L. & Cameron, A. G. W. The origin of the Moon and the single impact hypothesis. I. *Icarus* **66**, 515–535 (1986).
5. Benz, W., Cameron, A. G. W. & Melosh, H. J. The origin of the Moon and the single impact hypothesis. III. *Icarus* **81**, 113–131 (1989).
6. Cameron, A. G. W. & Benz, W. The origin of the Moon and the single impact hypothesis. IV. *Icarus* **92**, 204–216 (1991).
7. Cameron, A. G. W. The origin of the moon and the single impact hypothesis. V. *Icarus* **126**, 126–137 (1997).
8. Canup, R. M. & Esposito, L. W. Accretion of the Moon from an impact-generated disk. *Icarus* **119**, 427–446 (1996).
9. Ward, W. R. & Cameron, A. G. W. Disc evolution within the Roche limit. *Proc. Lunar Planet. Sci. Conf.* **9**, 1205–1207 (1978).
10. Stevenson, D. J. Origin of the moon—the collision hypothesis. *Annu. Rev. Earth Planet. Sci.* **15**, 271–315 (1987).
11. Canup, R. M. & Esposito, L. W. Accretion in the Roche zone: co-existence of rings and ringmoons. *Icarus* **113**, 331–352 (1995).
12. Goldreich, P. & Tremaine, S. The dynamics of planetary rings. *Annu. Rev. Astron. Astrophys.* **20**, 249–248 (1982).
13. Stewart, G. R. & Wetherill, G. W. Evolution of planetesimal velocities. *Icarus* **74**, 542–553 (1988).
14. Safronov, V. S. *Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets*. (Nauka, Moscow, 1969).
15. Lin, D. N. C. & Papaloizou, J. C. B. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 749–836 (Univ. Arizona Press, Tucson, 1993).
16. Makino, J. & Aarseth, S. J. On a Hermite integrator with Ahmad–Cohen scheme for gravitational many-body problems. *Publ. Astron. Soc. Jpn* **44**, 141–151 (1992).
17. Makino, J. A modified Aarseth code for GRAPE and vector processors. *Publ. Astron. Soc. Jpn* **43**, 859–876 (1991).

Acknowledgements. We thank A. G. W. Cameron and Jack Lissauer for their constructive criticisms, and W. Ward, L. Esposito and H. Tanaka for discussions and comments. S.I. thanks LASP, University of Colorado for hospitality. R.M.C. was supported by NASA's origins of solar systems programme, and G.R.S. was supported by NASA's planetary geology and geophysics programme.

Correspondence and requests for materials should be addressed to S.I. (e-mail: ida@geo.titech.ac.jp).

The optical counterpart of the isolated neutron star RX J185635–3754

Frederick M. Walter & Lynn D. Matthews

Department of Physics and Astronomy, Z=3800, State University of New York, Stony Brook, New York 11794-3800, USA

The extreme densities¹ of neutron stars make them an ideal system in which to investigate the equation of state of nuclear matter; accurate determinations of neutron star masses and radii are crucial for this. Current observations of neutron stars in binary systems yield masses that are generally consistent with theory². But measurements of radii are more difficult as they require the detection of thermal radiation from the surface, which in general is masked by emission from non-thermal processes in radio pulsars³ and X-ray binary systems⁴. Isolated radio-quiet neutron stars⁵ offer the best opportunity to observe the surface thermal emission. Here we report the detection of the optical counterpart of a candidate isolated neutron star, RX J185635–3754 (ref. 6). Our optical flux data, combined with existing extreme ultraviolet⁷ and X-ray⁶ observations, show the spectrum to be approximately thermal. By adopting the upper bound to the distance of the source, and assuming a plausible model for the spectral energy distribution, we find that the radius of the object cannot exceed 14 km. This result is inconsistent with a number of recent equations of state⁸ proposed for neutron stars.

RX J185635–3754 lies along a line of sight to the R CrA molecular cloud, which itself is only 130 pc away⁹; X-ray observations indicate that it must lie in front of the molecular cloud¹⁰, as the column of absorbing material is too small to include a molecular cloud. No optical counterpart had been found to a limiting magnitude of $V \approx 24$ (refs 6, 11, 12), which yields an X-ray-to-optical flux ratio $f_X/f_V > 7,000$. Such an extreme ratio rules out most objects other than neutron stars as possible sources⁶.

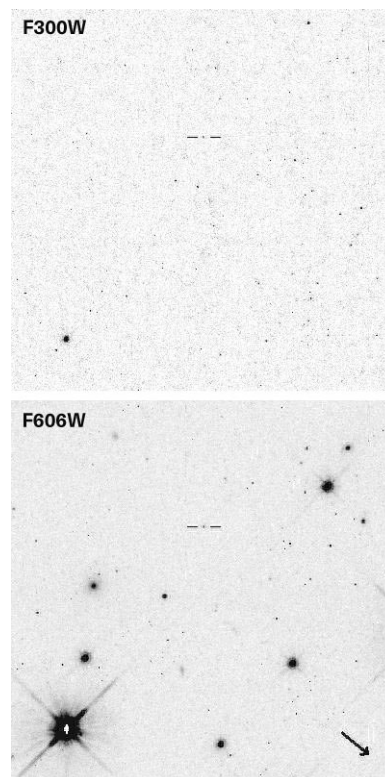
Figure 1 The central 400×400 pixels (17.5×17.5 arcsec) of the planetary camera images. The optical counterpart is marked. The arrow at the lower right of the F606W image points north. The bright star to the lower left is star F; star J is near the tail of the arrow. Both images are heavily stretched to bring out the faint features. The data were reduced using the WFPC2 team reduction method, which includes application of a high signal-to-noise superdark frame composed of dark frames taken as near as possible in time to the data of the observations, as well as an updated warm-pixel correction for highly time-dependent dark current by using a delta (local) dark frame²⁰. All images were processed by CR-SPLIT to facilitate cosmic-ray removal. The CR-SPLIT exposures were co-added using the VISTA PICCRS package. Final cosmic ray identification was done by blinking the split exposures in each filter. Residual cosmic rays were hand-cleaned only in the vicinity of the neutron star; they are evident elsewhere in the images. The dithered F606W frames were co-added using a simple adaptation of the Fruchter and Hook 'drizzling' algorithm²¹ written by A. M. Watson; we used 2×2 oversampling and a drop size of 1×1 oversampled pixels. F606W and F300W magnitudes of stars identified by Walter *et al.*⁶ (letters) and Campana *et al.*¹² (numbers) are: C, (–, 19.4); D, (–, 20.9); F, (–, 19.0); I, (18.7, 21.3); J, (20.8, 23.3); L, (–, 18.9); 21, (20.9, 24.6); 23, (23.6, 23.6); 24, (20.2, 22.9); 28, (21.4, >25). Stars C, D, F and L have saturated pixels in the F606W image.

The observations were made with the Hubble Space Telescope (HST) Wide Field and Planetary Camera 2 (WFPC2). They consist of a dithered pair of images through the F606W filter, totalling 4,400 s, and a single 2,400-s image in the F300W filter. The fully reduced images are shown in Fig. 1.

We identify the blue object at right ascension 18 h 56 min 35.41 s, declination $-37^\circ 54' 8''$ (J2000) as the optical counterpart. This position is referred to the reference frame of the Digital Sky Survey, by measuring the positions of stars in common between the two images. The offset (HST – DSS) is $+0.065$ s, $-1.1''$. The object lies 3.9 arcsec from the originally published X-ray position⁶, but only 2.0 arcsec from the revised position¹¹, which used a more accurate boresight correction. The formal uncertainty on the position is less than ± 0.02 arcsec, but we estimate that there may be uncertainties of up to 2.0 arcsec in the transformation to the DSS reference frame. The object has a stellar point spread function in each filter. The identification as the X-ray source counterpart is based on the blue colour of the object.

We determined the fluxes using aperture photometry. We measured the flux within an aperture of selectable size, and the background with an aperture of 40 pixels radius, with an inner radius 3 pixels outside the source extraction radius. The background was iteratively filtered to exclude points >3 standard deviations (s.d.) from the mean before computing the background level. We measured the source counts within a 9-pixel (0.41-arcsec) radius circle, and corrected to a 1-arcsec aperture using the measured point spread function of bright sources in the image. The flux in the F606W filter $f_{606} = 1.58 \pm 0.39 \times 10^{-19}$ erg cm⁻² s⁻¹ Å⁻¹; the flux in the F300W filter $f_{300} = 1.63 \pm 0.38 \times 10^{-18}$ erg cm⁻² s⁻¹ Å⁻¹. The main uncertainties in the fluxes are due to the Poisson noise in the background. The target is detected with a signal-to-noise ratio >8 in each band. On the Space Telescope (ST) magnitude system¹³, the f_{606} and f_{300} magnitudes are 25.9 and 23.4, respectively; f_{606} corresponds to a magnitude of 25.6 on the Vega magnitude system¹³, and f_X/f_V about 75,000. The $f_{300} - f_{606}$ colour is -2.5 on the ST magnitude system (an infinitely hot black body has an $f_{300} - f_{606}$ colour of -3.0 in this system).

The magnitudes are consistent with the identification as a



neutron star. The hot white dwarf Hz43 has about three times the X-ray flux of RX J185635–3754 and a comparably blue colour ($U - V = -1.2$), but at $V = 12.8$, f_U/f_V is close to unity. A lightly reddened hot white dwarf at $V = 25.6$ would lie at a distance of about 50 kpc. The counterpart must be much smaller than a white dwarf if it is foreground to the molecular cloud.

Little, if any, of the ultraviolet flux can be due to a red leak in the F300W filter. The reddest star, number 21 (ref. 12), has $f_{300}/f_{606} = 0.04$. We take this as the maximum possible red leak. As the neutron star candidate is considerably bluer than all the other stars, with $f_{300}/f_{606} = 10$, any red leak must be negligible.

If the object is a hot black body, the intrinsic f_{300}/f_{606} flux ratio is the ratio of the fourth power of the wavelengths, and the visual extinction $A_V = 0.52^{+0.13}_{-0.20}$ mag, using the Seaton¹⁴ extinction curve with $R = 3.1$. This is 2.3σ deviant from the measured X-ray absorption column of $1.4 \pm 0.1 \times 10^{20} \text{ cm}^{-2}$ ($A_V = 0.07$ mag), and might suggest the existence of spectral features in the optical spectrum. We note that excess emission near 6,000 Å is seen in the spectrum of isolated pulsars^{15,16}, but at much larger contrast. This emission could be attributable either to non-thermal emission from a magnetosphere or line emission from a surrounding nebula¹⁷; spectroscopic observations will be needed to characterize this excess.

The spectral energy distribution (SED) is shown in Fig. 2. We utilize the Rosat Position Sensitive Proportional Counter (PSPC) and High Resolution Imager data⁶, the EUVE Scanner count rate⁷, and the two optical points. The black-body fit to the X-ray data underpredicts the optical flux by a factor of 2.4 ± 0.5 at 3,000 Å and a factor of 3.7 ± 0.9 at 6,060 Å. This is evidence that the spectrum is more complex than a simple black body. Models with two black-body components, or more sophisticated models with surface temperature variations¹⁸, can match the f_{300} flux. Models involving radiative transfer through H, He or non-magnetic Fe atmospheres¹⁹ fail to predict the optical fluxes.

The blue colour and the reasonably close agreement of the optical fluxes with the extrapolated X-ray flux confirms that this object is the optical counterpart of the X-ray source. The isolated nature of the object, the approximately thermal character of its emission and the lack of strong variability on any timescale suggest that we are

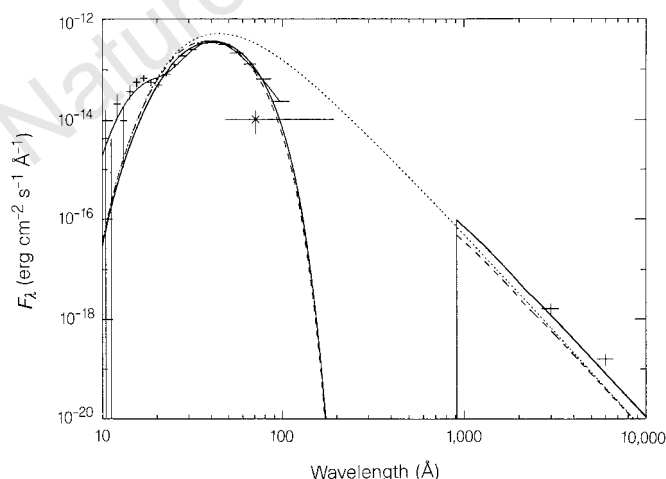


Figure 2 The spectral energy distribution of RX J185635–3754. The Rosat PSPC data⁶, the EUVE scanner flux⁷ and the two fluxes from the HST are plotted with error bars. The thin line is the black-body fit to the Rosat PSPC data, including the broadening effects of the instrumental response function. The dotted line is the unreddened 57 eV black body; the broken line is the black-body fit reddened by $A_V = 0.07$ mag (the gap between ~ 200 and 912 Å is due to absorption in the interstellar medium). The thick solid line is a reddened model with a sinusoidal temperature variation over the surface¹⁸ ($T_{\text{pole}} = 62$ eV; $T_{\text{equator}} = 0$). The f_{300} and f_{606} fluxes exceed those predicted by these simple models.

seeing the surface emission from a nearby neutron star. This is the only non-pulsating, single neutron star confirmed so far. As such, this object affords a unique opportunity to study the surface of a neutron star, and to study the interior equation of state. If the emission from the surface is truly thermal, then one can determine the angular diameter of the star. The radius follows from the parallax. Not knowing the distance, we can estimate upper limits to the radius by assuming that the neutron star is foreground to the R CrA molecular cloud.

The $kT_{\infty} = 57$ eV X-ray black body⁶ yields a radius as seen at infinity, $R_{\infty} < 8$ km. The models with more complex surface temperature distributions yield larger values for the radius. These models have more free parameters, including the inclinations of the rotation and magnetic poles to the line of sight and the functional form of the surface temperature distribution. A full analysis of these models is under way (P. An, J. M. Lattimer, F. M. W. and M. Prakash, manuscript in preparation). Here we illustrate that a simple model with the magnetic and rotation axes approximately co-aligned and in the plane of the sky (consistent with the lack of strong modulation of the X-ray flux) can match the observed SED. The temperature is assumed to vary as $(\text{latitude})^{0.25}$ (ref. 18). The polar cap temperature is 62 eV. This model SED, shown in Fig. 2, has $R_{\infty} < 14$ km, for a distance < 130 pc.

Although RX J185635–3754 offers the cleanest example now known of the surface of an isolated neutron star, there are significant unknowns. Although the upper limit to the distance is reasonably certain, the true distance is not known. Planned observations with the HST will permit us to measure the parallax and determine a firm lower limit of $R_{\infty} > 8(D/130 \text{ pc})$ km, where D is the distance in parsecs. Estimates of the radius are model-dependent and hence require a better understanding of the SED at all wavelengths. Planned X-ray and optical observations will define the SED better, and will indicate how much, if any, of the optical excess above a black body can be due to non-thermal emission processes. Assuming the latitudinally varying surface temperature model¹⁸, we find $R_{\infty} = 14(D/130 \text{ pc})$ km. This permits us to exclude most conventional equations of state⁸ (for mass $> 0.5 M_{\odot}$) if $D < 100$ pc. Similarly, for mass $> 1 M_{\odot}$, we can rule out equations of state incorporating kaon condensates⁸ if D is less than ~ 75 pc. We expect that further observations of RX J185635–3754 will yield important insights into the nature of the surfaces and the interior equation of state of neutron stars. □

Received 7 April; accepted 24 July 1997.

- Baym, G. & Pethick, C. Physics of neutron stars. *Annu. Rev. Astron. Astrophys.* **17**, 415–443 (1979).
- van Paradijs, J. & McClintock, J. E. in *X-Ray Binaries* (eds Lewin, W. H. G., van Paradijs, J. & van den Heuvel, E. P. J.) 58–125 (Cambridge Univ. Press, 1995).
- Phinney, E. S. & Kulkarni, S. R. Binary and millisecond pulsars. *Annu. Rev. Astron. Astrophys.* **32**, 591–639 (1994).
- White, N. E., Nagase, F. & Parmar, A. N. in *X-Ray Binaries* (eds Lewin, W. H. G., van Paradijs, J. & van den Heuvel, E. P. J.) 1–57 (Cambridge Univ. Press, 1995).
- Caraveo, P. A., Bignami, G. F. & Trümper, J. Radio-silent isolated neutron stars as a new astronomical reality. *Astron. Astrophys. Rev.* **7**, 209–216 (1996).
- Walter, F. M., Wolk, S. J. & Neuhäuser, R. Discovery of a nearby isolated neutron star. *Nature* **379**, 233–235 (1996).
- Lampton, M. et al. An all-sky catalog of faint Extreme Ultraviolet sources. *Astrophys. J. Suppl.* **108**, 545–557 (1997).
- Prakash, M. & Lattimer, J. M. The radius of the neutron star RX J185635–3754 and implications for the equation of state. *Astrophys. J. Lett.* (submitted).
- Marraco, H. G. & Rydgren, A. E. On the distance and membership of the R CrA association. *Astron. J.* **86**, 62–68 (1981).
- Wang, Q. D. in *The Soft X-Ray Cosmos* (AIP Conference Proceedings 313) (eds Schlegel, E. M. & Petre, R.) 301–305 (AIP, New York, 1994).
- Neuhäuser, R., Thomas, H.-C., Danner, R., Peschke, S. & Walter, F. M. On the X-ray position and deep optical imaging of the neutron star candidate RXJ1856.5–3754. *Astron. Astrophys.* **318**, L43–L46 (1996).
- Campana, S., Mereghetti, S. & Sidoli, L. Is RXJ1856.5–3754 an old neutron star? *Astron. Astrophys.* **320**, 783–786 (1997).
- Leitherer, C. (ed.) *HST Data Handbook* (Space Telescope Science Inst., Baltimore, 1995).
- Seaton, M. Interstellar extinction in the UV. *Mon. Not. R. Astron. Soc.* **187**, 73P–76P (1979).
- Bignami, G. F., Caraveo, P. A., Mignani, R., Edelstein, J. & Bowyer, S. Multiwavelength data suggest cyclotron feature on the hot thermal continuum of Geminga. *Astrophys. J. Lett.* **456**, L111–L114 (1996).
- Pavlov, G. G., Stringfellow, G. S. & Córdoba, F. A. Hubble Space Telescope observations of pulsars. *Astrophys. J.* **467**, 370–384 (1996).
- Blaes, O., Warren, O. & Madau, P. Accreting, isolated neutron stars: III. Preheating of infalling gas and cometary H II regions. *Astrophys. J.* **454**, 370–381 (1995).

18. Greenstein, G. & Hartke, G. J. Pulselike character of blackbody radiation from neutron stars. *Astrophys. J.* **271**, 283–293 (1983).
19. Pavlov, G. G., Zavlin, V., Trümper, J. & Neuhäuser, R. Multiwavelength observations of isolated neutron stars as a tool to prove the properties of their surfaces. *Astrophys. J. Lett.* **472**, L33–L36 (1996).
20. Holtzman, J. A. *et al.* The photometric performance and calibration of WFPC2. *Pub. Astron. Soc. Pacif.* **107**, 1065–1093 (1995).
21. Fruchter, A. S. & Hook, R. N. in *The Hubble Space Telescope and the High Redshift Universe* (eds Tanvir, N. R., Salamanca, A. & Wall, J. V.) (World Scientific, Singapore, 1997).

Acknowledgements. We thank A. M. Watson for performing the bulk of the image reduction, and J. Holtzman for preparing and providing the delta dark and superdark frames necessary for our data; we also thank them both for their time and patience in the discussion of data reduction issues. We acknowledge useful discussions with P. An, J. Lattimer and R. Neuhäuser.

Correspondence and requests for materials should be addressed to F.M.W. (e-mail: fwalter@astro.sunysb.edu).

A general boundary condition for liquid flow at solid surfaces

Peter A. Thompson* & Sandra M. Troian†

* The Celerity Group, 20 Nassau Street, Suite 209, Princeton, New Jersey 08542, USA

† Department of Chemical Engineering, Princeton University, Princeton, New Jersey 08544, USA

Modelling fluid flows past a surface is a general problem in science and engineering, and requires some assumption about the nature of the fluid motion (the boundary condition) at the solid interface. One of the simplest boundary conditions is the no-slip condition^{1,2}, which dictates that a liquid element adjacent to the surface assumes the velocity of the surface. Although this condition has been remarkably successful in reproducing the characteristics of many types of flow, there exist situations in which it leads to singular or unrealistic behaviour—for example, the spreading of a liquid on a solid substrate^{3–8}, corner flow^{9,10} and the extrusion of polymer melts from a capillary tube^{11–13}. Numerous boundary conditions that allow for finite slip at the solid interface have been used to rectify these difficulties^{4,5,11,13,14}. But these phenomenological models fail to provide a universal picture of the momentum transport that occurs at liquid/solid interfaces. Here we present results from molecular dynamics simulations of newtonian liquids under shear which indicate that there exists a general nonlinear relationship between the amount of slip and the local shear rate at a solid surface. The boundary condition is controlled by the extent to which the liquid ‘feels’ corrugations in the surface energy of the solid (owing in the present case to the atomic close-packing). Our generalized boundary condition allows us to relate the degree of slip to the underlying static properties and dynamic interactions of the walls and the fluid.

During the past decade molecular dynamics simulations have emerged as a powerful tool for probing the microscopic behaviour of liquids at interfaces (ref. 15, and references therein). Studies have demonstrated how solids induce order in adjacent liquids and how this order, in turn, controls the amount of momentum transfer¹⁶. The no-slip condition has been shown to be just one of many allowable flow boundary conditions ranging from pure slip to multi-layer locking. The degree of slip at the boundary depends on a number of interfacial parameters including the strength of the liquid–solid coupling, the thermal roughness of the interface, and the commensurability of wall and liquid densities¹⁶. These findings have led to a new understanding of stick-slip phenomena in boundary lubrication¹⁷ and have revealed the sensitivity of liquid spreading to microstructure at the liquid/solid interface^{7,8}. Earlier work¹⁶ investigated flows in a regime for which the degree of slip was independent of local shear rate $\dot{\gamma}$. This regime is described by the well-known linear Navier boundary condition¹ in which $\Delta V = L_s \dot{\gamma}$, where ΔV is the velocity difference between the solid and adjacent liquid, and L_s is a constant slip length. Here we examine the

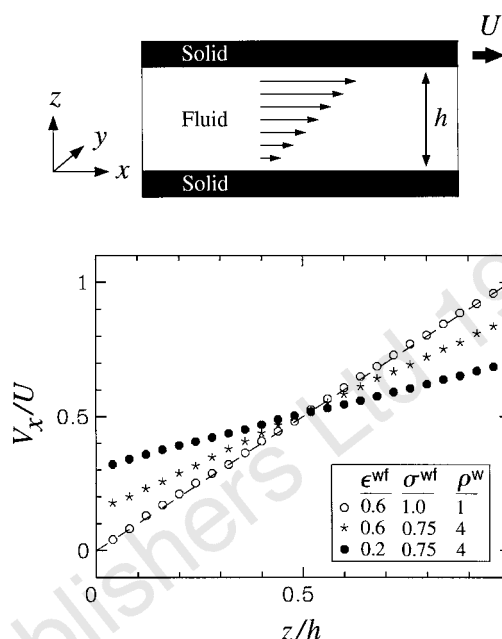


Figure 1 Steady-state flow profiles and schematic of the Couette flow geometry. The Couette cell measured $12.51\sigma \times 7.22\sigma \times h$ where h varied from 16.71σ to 24.57σ . The number of fluid molecules ranged from 1,152 to 1,728, respectively. The $\hat{x}$ -direction of the cell is aligned along the $[11\bar{2}]$ orientation of the face-centred cubic lattice comprising the wall, and periodic boundary conditions are imposed along $\hat{x}$ and $\hat{y}$. The flow profiles were obtained for systems with $U = 1.0\sigma\tau^{-1}$, $h = 24.57\sigma$, and walls characterized by the indicated density and Lennard–Jones parameters. Values for ϵ^{wf} , σ^{wf} and ρ^w (see text) are in units of ϵ , σ and ρ , respectively. Following an equilibration period of $\sim 100\tau$, the profiles were computed by averaging the instantaneous particle velocities within bins of width $\sim 1\sigma$ spanning the distance between the two walls. The duration of the averaging varied from 250τ to $7,000\tau$ depending on the signal-to-noise ratio. Accurate resolution of flows with $\dot{\gamma} < 0.01\tau^{-1}$ typically required $> 2,500\tau$ of averaging. The dashed line indicates Couette flow with a no-slip boundary condition.

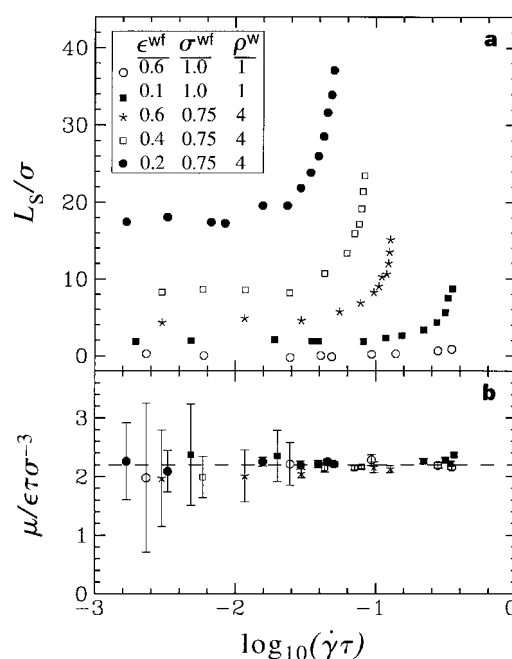


Figure 2 Variation of the slip length L_s (panel **a**) and viscosity μ (panel **b**) as a function of shear rate for systems with the indicated interfacial properties. L_s was computed from the definition $L_s = \Delta V_x / \dot{\gamma}$, which for Couette flow reduces to $(U/h - \dot{\gamma})/2$; μ was computed from the relation $\mu = P_{zx} / \dot{\gamma}$, where P_{zx} is the zx -component of the microscopic stress tensor averaged across the cell¹⁹.

boundary condition over a much wider range of shear rates which reveals a nonlinear relationship between L_s and $\dot{\gamma}$.

Our molecular dynamics simulations model a simple liquid undergoing planar shear in a Couette cell as shown in Fig. 1. Each wall of the cell consists of N_w atoms forming two (111) planes of a face-centred cubic lattice. N_w ranges from 144 to 576 depending on the wall density, ρ^w . The liquid is treated as an isothermal ensemble of spherical molecules. Molecules separated by a distance r interact through a shifted Lennard–Jones 6–12 potential:

$$V^{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 - \left(\frac{\sigma}{r_c} \right)^{12} + \left(\frac{\sigma}{r_c} \right)^6 \right] \quad (1)$$

where ϵ and σ are characteristic energy and length scales, and the potential is zero for $r > r_c = 2.2\sigma$. The wall–fluid interaction is also modelled with a truncated Lennard–Jones potential with energy and length scales ϵ^{wf} and σ^{wf} , respectively, and cut-off r_c . Similar systems have been widely used to investigate a variety of bulk and interfacial flows¹⁵.

The equilibrium state of the fluid is a well-defined liquid phase characterized by number density $\rho = 0.81\sigma^{-3}$ and temperature $T = 1.1k_B/\epsilon$. Constant temperature is maintained by weakly coupling the particle dynamics to a thermal reservoir through the addition of Langevin noise and frictional terms to the equations of motion^{7,8,16–18}. Good thermal coupling obviates the need for the noise and frictional terms in the $\hat{x}$ and $\hat{z}$ directions^{16,18}. The resulting equation of motion for the y degree of freedom of the i th molecule of mass m is

$$m\ddot{y}_i = \sum_{j \neq i} \partial V^{LJ} / \partial y_i - m\Gamma \dot{y}_i + \eta_i \quad (2)$$

where $\sum_{j \neq i}$ denotes a sum over all interactions with i , Γ is a friction constant that controls the rate of heat exchange with the reservoir, and η_i is a gaussian distributed random force with zero mean and variance $2mk_B T \Gamma$. Efficient temperature control with minimal effect on particle dynamics was achieved with $\Gamma = 1.0\tau^{-1}$, where $\tau = (m\sigma^2/\epsilon)^{1/2}$ is the characteristic time of the Lennard–Jones potential. The equations of motion are integrated using a fifth-order Gear predictor–corrector algorithm¹⁹ with a time step $\Delta t = 0.005\tau$.

The simulations produce steady-state velocity fields as shown in Fig. 1. The flows were induced by translating the upper wall with velocity U in the $\hat{x}$ direction. The profiles recover the expected flow behaviour from continuum hydrodynamics with boundary conditions involving varying degrees of slip. For the system with $\rho^w/\rho = 1$ and $\epsilon^{wf}/\epsilon = 0.6$, $\Delta V_x/U = 0$ and the no-slip condition is obeyed exactly. For larger ρ^w or smaller ϵ^{wf} , slip is evident. This variation with wall properties is consistent with that reported earlier¹⁶—namely, the amount of momentum transfer at the wall/fluid interface decreases as the relative surface energy corrugation of the wall decreases. The corrugation is maximized when the wall and fluid densities are commensurate ($\rho^w = \rho$) and the strength of the wall–fluid coupling ϵ^{wf} is large. In this case, there is efficient momentum transfer and the resulting flow is consistent with a no-slip, or in extreme cases, a “stick” boundary condition¹⁶. For incommensurate densities or smaller couplings, the corrugation is weaker and interfacial slip develops.

We now consider how the fluid responds to variations in the wall speed U for five different sets of interfacial parameters (ϵ^{wf} , σ^{wf} and ρ^w). To generalize the analysis, the results are expressed in terms of the shear rate $\dot{\gamma}$ within the fluid measured directly from the computed linear velocity profiles as in Fig. 1. Note that the bulk viscosity μ of the fluid is a constant over the entire range of shear rates (Fig. 2b) indicating the bulk fluid remains newtonian. Based on previous investigations of shear-induced ordering in simple fluids, non-newtonian response of the bulk fluid is expected for $\dot{\gamma} \geq 2\tau^{-1}$ (ref. 20).

The variation in L_s as a function of $\dot{\gamma}$ is shown in Fig. 2a. The data

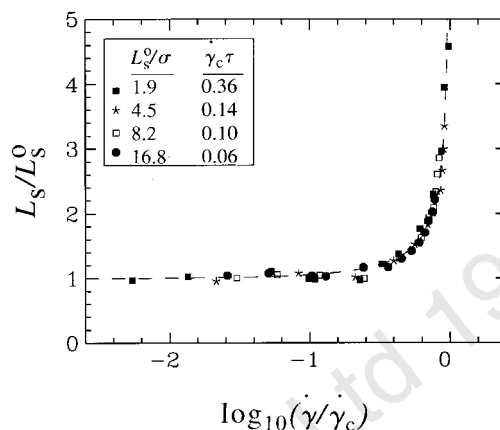


Figure 3 Master curve describing the flow boundary condition. The data is the same as that shown in Fig. 2a with L_s and $\dot{\gamma}$ scaled by the indicated values of L_s^0 and $\dot{\gamma}_c$. The dashed line represents $L_s = L_s^0(1 - \dot{\gamma}/\dot{\gamma}_c)^{-1/2}$.

reveal several significant features. At sufficiently low $\dot{\gamma}$, the boundary condition is consistent with the Navier model. In this regime, the slip length equals its limiting value L_s^0 which ranges from zero to $\sim 17\sigma$ for the interfacial parameters chosen. In general, the amount of slip increases with decreasing surface energy corrugation. The latter was achieved by either decreasing the wall–fluid coupling ϵ^{wf} or increasing the wall density ρ^w . At higher shear rates, the Navier condition breaks down as the slip length increases rapidly with $\dot{\gamma}$. Surprisingly, the boundary condition is nonlinear even though the liquid is still newtonian. In fact the slip length appears to diverge as the shear rate approaches a critical value $\dot{\gamma}_c$. This value depends on the surface energy corrugation: as the corrugation decreases, so too does $\dot{\gamma}_c$. We note that the values for $\dot{\gamma}_c$ shown in Fig. 2 are of the same magnitude as the shear rates previously found in molecular dynamics studies of low capillary number flows^{7,8}.

The functional behaviour for L_s suggests a universal boundary condition at a solid–liquid interface. Scaling L_s by its asymptotic limiting value L_s^0 and $\dot{\gamma}$ by its critical value $\dot{\gamma}_c$ collapses the data onto the curve shown in Fig. 3. This curve is well described by the form $L_s = L_s^0(1 - \dot{\gamma}/\dot{\gamma}_c)^{-\alpha}$ where the dashed line represents the value $\alpha = 1/2$. This behaviour suggests that close to a critical shear rate the boundary condition can significantly affect flow behaviour at macroscopic distances from the wall, an experimental observation repeatedly seen in many polymeric systems²¹. Such significant slip is not easily recoverable in the phenomenological models presented to date. It also suggests that for flows in the vicinity of $\dot{\gamma}_c$ small changes in the surface properties can lead to large fluctuations in the apparent boundary condition.

Given the remarkable collapse of the data, one would like to understand what causes L_s to diverge at $\dot{\gamma}_c$ and what microscopic properties determine its scaling behaviour. To understand this divergence, we have considered the much simpler problem of a single particle undergoing Langevin dynamics in a driven periodic potential. (Independently, a similar model has been used to study stick-slip dynamics in boundary lubrication²².) In this model, a particle is held at a fixed height above a solid surface which is translating uniformly with speed v . The surface, composed of Lennard–Jones atoms, is analogous to that used in the multi-particle molecular dynamics simulations. As in the collective system, the motion of the particle is independent of the wall velocity at low v and experiences significant slip at high v . The particle achieves pure slip for velocities $v > v_c$. The less corrugated the surface potential, the smaller v_c becomes and the larger the slip at low velocities.

It is natural to assume that the scaling for v_c (or equivalently $\dot{\gamma}_c$)

is set by the liquid–solid interaction timescale, namely $\gamma_c^{-1} \sim \tau^{wf} = (m\sigma^{wf}/\epsilon^{wf})^{1/2}$. According to this scaling, increasing ρ^{wf} (or likewise decreasing σ^{wf}) should lead to larger values of γ_c . This behaviour is not supported by the data shown in Fig. 2 in which smaller values of σ^{wf} lead to smaller γ_c . We therefore propose a different scaling parameter more closely associated with the surface energy corrugation such that it vanishes in the limit of small ϵ^{wf} and large ρ^{wf} . For both these limits, we expect L_s to increase and $\gamma_c \rightarrow 0$. In previous interfacial studies, geometric measures of real surface roughness, like the areal projection of a rough terrain, have been used as a key microscopic parameter controlling slip⁴. We introduce instead a roughness parameter R which characterizes the coarseness of the potential surface, $\phi(\bar{r})$, measured at some fixed height above the wall by a test particle. This potential surface is represented by the vector $x\hat{x} + y\hat{y} + \phi^{wf}/\epsilon^{wf}\phi(r)\hat{z}$. With this definition $R = A/A_\infty - 1$ where $A = \int [1 + (\partial\phi/\partial x)^2 + (\partial\phi/\partial y)^2]^{1/2} dx dy$, $A_\infty = \int dx dy$, $\phi(\bar{r}) = \Sigma_{N_w} V^{LJ}(|\bar{r} - \bar{r}_w^i|)$, $\bar{r}_w^i$ denotes the position of a wall atom, and $\bar{r}$ locates the test particle.

For the single-particle model just described, v_c scales as $R^{1/2}$ over a wide range of interfacial parameters. This correlation suggests the possibility of describing slip behaviour in terms of a single scalar quantity. But extensions of this analysis to a collective interacting system are complicated by the possibility of layering phenomena at the wall–fluid interface, the dimensionality of the full system, and inertial corrections. Nonetheless, for the limited set of data shown in Fig. 2, we find that γ_c scales as a power in R with an exponent close to 3/4. Although more studies of this type are required to predict the precise relationship between γ_c and R , it is possible that a scalar measure of the roughness of the energy surface is a key parameter controlling the amount of slip experienced by a liquid under shear.

Our results indicate that the well-known Navier slip boundary condition is but the low-shear-rate limit of a more generalized universal relationship which is significantly nonlinear and divergent at a critical shear rate, γ_c . The value of γ_c is set by the corrugation of the potential surface which signals the point at which the solid can no longer impart momentum to the liquid. This means that the same liquid molecules sheared against different substrates will experience varying amounts of slip and vice versa. The functional dependence for L_s shown in Fig. 2a provides a new non-phenomenological boundary condition that can be used to model viscous flows along solid surfaces. This relation provides a mechanism for relieving the well-known stress singularity in spreading liquids and corner flows as it naturally allows for varying degrees of slip on approach to regions of higher shear stress and shear rate. □

Received 23 May; accepted 8 July 1997.

1. Lamb, H. *Hydrodynamics* (Dover, New York, 1932).
2. Batchelor, G. K. *An Introduction to Fluid Dynamics* (Cambridge Univ. Press, 1967).
3. Huh, C. & Scriven, L. E. Hydrodynamic model of steady movement of a solid/liquid/fluid contact line. *J. Colloid. Interface. Sci.* **35**, 85–101 (1971).
4. Hocking, L. M. A moving fluid interface on a rough surface. *J. Fluid Mech.* **76**, 801–817 (1976).
5. Dussan, E. B. On the spreading of liquids on solid surfaces: static and dynamic contact lines. *Annu. Rev. Fluid. Mech.* **11**, 371–400 (1979).
6. Koplik, J., Banavar, J. R. & Willemsen, J. F. Molecular dynamics of Poiseuille flow and moving contact lines. *Phys. Rev. Lett.* **60**, 1282–1285 (1988); Molecular dynamics of fluid flow at solid surfaces. *Phys. Fluids A* **1**, 781–794 (1989).
7. Thompson, P. A. & Robbins, M. O. Simulations of contact line motion: slip and the dynamic contact angle. *Phys. Rev. Lett.* **63**, 766–769 (1989).
8. Thompson, P. A., Brinkerhoff, W. B. & Robbins, M. O. Microscopic studies of static and dynamic contact angles. *J. Adhesion Sci. Technol.* **7**, 535–554 (1993).
9. Moffatt, H. K. Viscous and resistive eddies near a sharp corner. *J. Fluid Mech.* **18**, 1–18 (1964).
10. Koplik, J. & Banavar, J. R. Corner flow in the sliding plate problem. *Phys. Fluids* **7**, 3118–3125 (1995).
11. Pearson, J. R. A. & Petrie, C. J. S. in *Polymer Systems: Deformation and Flow* (eds Wetton, R. E. & Whorlow, R. W.) 163–187 (Macmillan, London, 1968).
12. Richardson, S. On the no-slip boundary condition. *J. Fluid Mech.* **59**, 707–719 (1973).
13. Denn, L. M. Issues in viscoelastic fluid mechanics. *Annu. Rev. Fluid Mech.* **22**, 13–34 (1990).
14. Durbin, P. A. Considerations on the moving contact-line singularity, with application to frictional drag on a slender drop. *J. Fluid Mech.* **197**, 157–169 (1988).
15. Koplik, J. & Banavar, J. R. Continuum deductions from molecular hydrodynamics. *Annu. Rev. Fluid Mech.* **27**, 257–292 (1995).
16. Thompson, P. A. & Robbins, M. O. Shear flow near solids: epitaxial order and flow boundary conditions. *Phys. Rev. A* **41**, 6830–6837 (1990).
17. Thompson, P. A. & Robbins, M. O. The origin of stick-slip motion in boundary lubrication. *Science* **250**, 792–794 (1990).
18. Stevens, M. J. *et al.* Comparison of shear flow of hexadecane in a confined geometry and in bulk. *J. Chem. Phys.* **106**, 7303–7314 (1997).

19. Allen, M. & Tildesley, D. *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
20. Loose, W. & Hess, S. Rheology of dense fluids via nonequilibrium molecular hydrodynamics: shear thinning and ordering transition. *Rheol. Acta.* **28**, 91–101 (1989).
21. Atwood, B. T. & Schowalter, W. R. Measurements of slip at the wall during flow of high-density polyethylene through a rectangular conduit. *Rheol. Acta.* **28**, 134–146 (1989).
22. Rozman, M. G., Urbakh, M. & Klafter, J. Stick-slip motion and force fluctuations in a driven two-wave potential. *Phys. Rev. Lett.* **77**, 683–686 (1996).

Acknowledgements. P.A.T. thanks the Exxon Education Foundation and an NSF CAREER award which helped initiate these studies. S.M.T. was supported by the Exxon Education Foundation and the NCI through a Research Initiation and CAREER award, and a seed grant from the MRSEC program of the Princeton Materials Institute.

Correspondence should be addressed to S.M.T. (e-mail: stroian@princeton.edu).

Laser action in organic semiconductor waveguide and double-heterostructure devices

V. G. Kozlov, V. Bulović, P. E. Burrows & S. R. Forrest

Center for Photonics and Optoelectronic Materials, Department of Electrical Engineering and the Princeton Material Institute, Princeton University, Princeton, New Jersey 08544, USA

Stimulated emission by optical pumping of solid-state organic materials has been well known since the late 1960s following the first demonstrations of laser action in dye-doped gels and molecular crystals^{1–4}. Interest in this field has been revived by the demonstration of efficient, long-lived and intense electroluminescence in both polymeric⁵ and small-molecular-weight⁶ organic thin films, which indicates the possibility of laser action in these materials. Several recent studies of optically pumped polymers have reported emission phenomena suggestive of laser action^{7–9}. Here we present clear evidence for laser action from optically pumped, vacuum-deposited thin films of organic molecules, in both slab-waveguide and double-heterostructure configurations. This realization of laser action in conducting organic thin films should open the way to the development of a new class of electrically pumped laser diodes.

Lasing (in contrast to amplified spontaneous emission, superluminescence or related phenomena) can be unambiguously identified from five phenomena: (1) a clear indication of a threshold in output energy as a function of input (or pump) energy, with a high lasing efficiency above threshold; (2) strong output beam polarization; (3) spatial coherence (as indicated by a diffraction-limited output beam or speckle); (4) significant spectral line narrowing; and (5) the existence of laser cavity resonances, or modes. Although several recent reports^{7–9} on polymer-based optically pumped thin films have discussed structures showing one or two of the above phenomena, there has (to our knowledge) been no report of all of these properties having been observed in such a system.

Tris-(8-hydroxyquinoline) aluminum (Alq₃) doped with 2.5% DCM laser dye provides an excellent laser material¹⁰ as the red stimulated emission of DCM at a wavelength of $\lambda = 645$ nm is far from the ultraviolet absorption edge of the Alq₃ host (at $\lambda = 450$ nm)¹¹, whereas the absorbance of the DCM is centred near the emission maximum of Alq₃ (at $\lambda = 530$ nm), thereby providing for efficient Förster energy transfer¹² from ultraviolet-excited Alq₃ (see Fig. 1 for molecular structural formulae of Alq₃ and DCM). Doping also allows for reduction of the density of the optically active DCM molecules (thereby reducing the effective density of states), which lowers the threshold and increases the efficiency of a laser¹³. In fact, efficient orange electroluminescence for DCM-doped Alq₃ films has been demonstrated⁶.

Lasers shown in Fig. 1 were grown on InP substrates pre-coated

with a 2- μm -thick layer of SiO_2 , deposited by plasma-enhanced chemical vapour deposition. The organic films were deposited by high vacuum (5×10^{-7} torr) co-evaporation of 40 : 1 (mass ratio) Alq_3 and DCM. A 3,000- $\text{\AA}$ -thick film of Alq_3/DCM (with optical index of refraction $n = 1.7$) forms a slab optical waveguide, with SiO_2 ($n = 1.4$) as a cladding layer on one side, and air ($n = 1$) on the other. In a separate experiment, a 500- $\text{\AA}$ -thick film of Alq_3/DCM was sandwiched between two 1,250- $\text{\AA}$ -thick Alq_3 cladding layers to form an in-plane optical waveguide. This double heterostructure uses all conducting organic layers, and has the properties of both electrical and optical confinement which are required to achieve low-threshold lasing under electrical injection¹⁴. Before organic film deposition, the (100)InP substrate was cleaved, ultimately defining the laser length while forming parallel edges. The vacuum-deposited organic film conforms to the shape of the underlying substrate, resulting in optically smooth, parallel facets with reflectivities $\sim 7\%$.

Lateral confinement of the optical mode in both structures is achieved by gain-guiding induced by the optical pump beam (gain-guiding is the process in which the optical mode is confined in regions where the optical gain is at a maximum). The pump beam was generated with a pulsed nitrogen laser beam ($\lambda = 337$ nm; pulse width, 500 ps at a 50-Hz repetition rate) and focused by a cylindrical lens into a 50- μm -wide stripe on the film surface orientated orthogonal to the edges (and hence the facets) of the film (Fig. 1a). A photograph of an operating slab laser device (Fig. 1b) shows a well defined beam of bright red laser emission in a direction orthogonal to the facets. Examination of the output emission

revealed a nearly diffraction limited beam in the direction of the substrate normal, with speckle clearly apparent in the far-field pattern. In the in-plane direction, the beam width was somewhat less than that of the pump beam, and was bounded by an orange 'halo' characteristic of the spontaneous emission spectrum of DCM⁶.

Figure 2 shows the edge emission spectra for a 5-mm-long double-heterostructure laser taken near threshold using a spectrograph with a CCD (charge-coupled device) camera. Below threshold (at a pump energy of $\sim 1.2 \mu\text{J cm}^{-2}$), the spectrum is dominated by a broad spontaneous emission background centred at $\lambda = 640$ nm. Strong laser emission accompanied by a dramatic decrease in spectral linewidth appears as a set of sharp peaks on the long-wavelength ($\lambda = 645$ nm) side of the spontaneous spectrum at pump energy densities $> 1.2 \mu\text{J cm}^{-2}$.

Figure 3 shows a high-resolution spectrum for a short-cavity, 500- μm -long slab laser, revealing numerous peaks (inset) evenly spaced at 2 $\text{\AA}$, corresponding to the free spectral range of the cavity, clearly indicating resonant cavity modes. The spectral width of the peaks (1 $\text{\AA}$ full-width at half-maximum) is limited by the resolution of our spectrograph. The laser emission is polarized in the plane of the organic film (TE mode), with the degree of polarization > 15 dB (Fig. 2 inset), also limited by our experimental resolution. The laser output power was highly stable—that is, no significant degradation in the laser performance was observed after operating for ~ 24 h at a very high peak power of ~ 50 W in a nitrogen atmosphere ($\sim 4 \times 10^6$ pulses). Such stability is essential for the eventual

Figure 1 Details of the lasers used in this work. **a**, Schematic of the laser structure and experimental set-up. **b**, Photograph of the Alq_3/DCM slab-waveguide laser showing the red output beam diverging with distance from the laser facets indicated by the intense radiation on the black InP substrate (photograph courtesy of A. Forrest). **c**, Chemical structural formulae of DCM and Alq_3 .

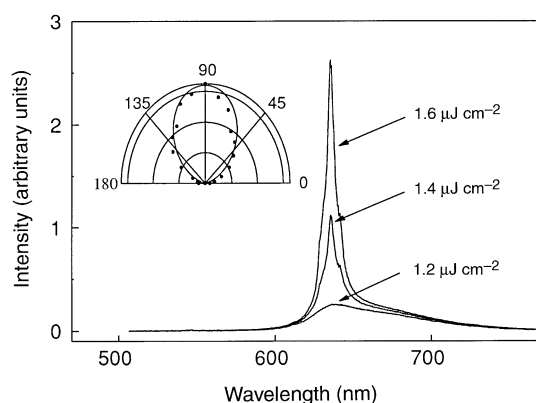
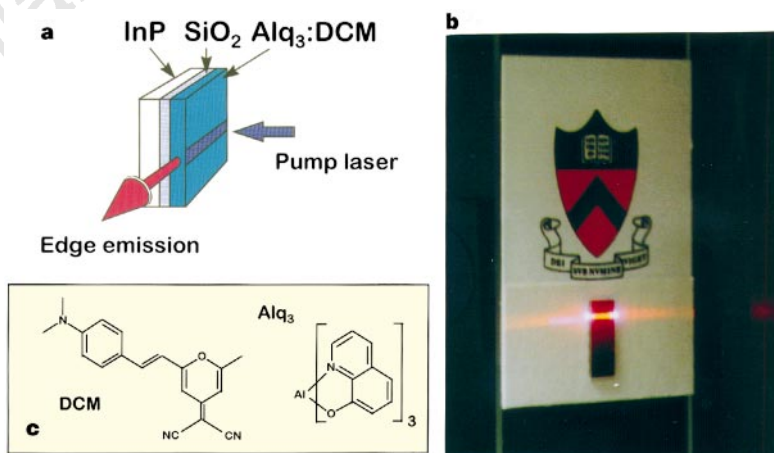


Figure 2 Spectra of the edge emission from an Alq_3/DCM slab laser at increasing excitation levels near threshold. Inset, polarization dependence of the laser emission as a function of angle (α) between a plane orthogonal to the film surface and the plane of the polarizer (data points: the solid line follows $\sin^2(\alpha)$, the expected dependence for TE polarization).

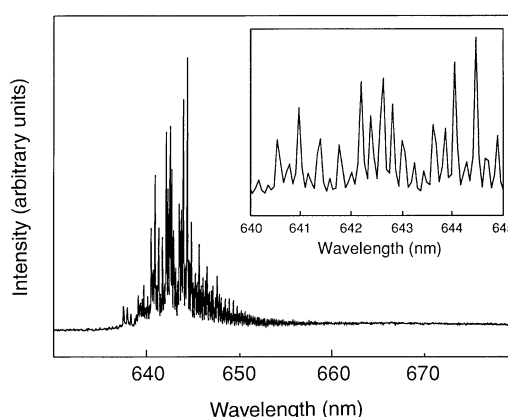


Figure 3 High-resolution edge emission spectrum of a 500- μm -long slab laser at a pump energy density of $1.7 \mu\text{J cm}^{-2}$. Inset, a portion of the same spectrum under higher resolution, showing evenly spaced resonant cavity modes and indicating a modal full-width at half-maximum of $< 1 \text{\AA}$.

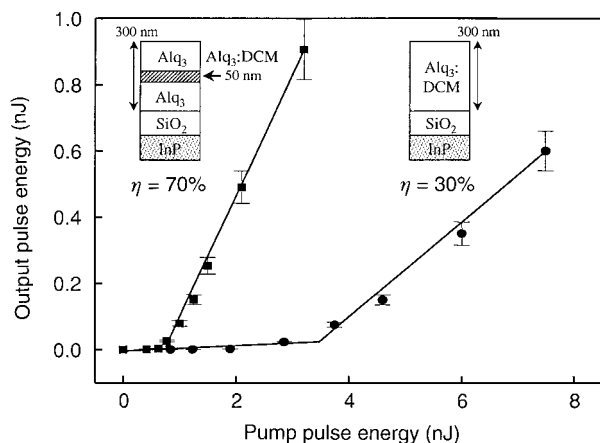


Figure 4 Dependence of output energy on the input pump energy near threshold for 1-mm-long slab-waveguide (right inset) and double-heterostructure (left inset) laser devices whose structures are illustrated. Solid lines are fits to the data; η is the laser slope efficiency.

practical application of electrical injection lasers based on these compounds.

The dependence of output energy on the pump energy for the slab and double-heterostructure devices, shown in Fig. 4, clearly indicates a threshold. To reduce mirror losses, long (~ 1 -mm) laser cavities were used to obtain the differential quantum efficiency (η) above threshold. We find that $\eta = 30\%$ for a slab device, and $\eta = 70\%$ for a double-heterostructure laser. The significantly higher differential efficiency and lower threshold of the double-heterostructure device as compared to the slab structure provides clear evidence for optical confinement in the Alq₃/DCM layer as compared to the cladding, as required for low-threshold electrically pumped injection lasers.

The thickness of the active layer in electrically pumped double heterostructures is ultimately limited by the hole diffusion length in Alq₃/DCM film, which is ~ 100 Å (ref. 6). To assess the prospects for lasing in such structures, we also fabricated a double-heterostructure device with a 50-Å-thick active layer, which had a threshold pump energy density of $1 \mu\text{J cm}^{-2}$ corresponding to a carrier density of $1.5 \times 10^{12} \text{ cm}^{-2}$. Assuming a spontaneous radiative lifetime of 10 ns for Alq₃ and DCM¹⁵, and assuming that only 25% of the electrically injected carriers form singlet excitons¹⁶, we estimate a threshold current density for pulsed electrical injection of $\sim 100 \text{ A cm}^{-2}$, which is achievable under low-duty-cycle pulsed excitation in electrically pumped organic light-emitting devices based on Alq₃. One obstacle which might prevent the realization of small-molecule-based organic lasers is non-radiative exciton recombination (or quenching) previously observed at high carrier densities. For example, bimolecular reactions in Alq₃ films¹⁵ have been found to decrease the photoluminescent quantum efficiency significantly at incident intensities above $10^{14} \text{ photons cm}^{-2}$. However, the lasing threshold in Alq₃/DCM films discussed here is at pump intensities of only $10^{12} \text{ photons cm}^{-2}$, leaving ample room for increased pump intensities without a decrease in efficiency.

We have shown low-threshold ($1 \mu\text{J cm}^{-2}$), high-efficiency (70%), narrow-linewidth (< 1 Å), high-output-power (> 50 W) lasing at $\lambda = 645$ nm in vacuum-deposited organic semiconductor thin films. The pump thresholds for organic double heterostructures are estimated at 100 A cm^{-2} for an electrically pumped laser. The ease of processing, low threshold, high stability and other characteristics of vacuum deposited materials should open the way towards realizing a new generation of electrically pumped solid-state lasers based on organic thin films. □

Received 29 April; accepted 15 August 1997.

1. Soffer, B. H. & McFarland, B. B. Continuously tunable, narrow-band organic dye lasers. *Appl. Phys. Lett.* **10**, 266–267 (1967).

- Kogelnik, H. & Shank, C. V. Stimulated emission in a periodic structure. *Appl. Phys. Lett.* **18**, 152–154 (1971).
- Karl, N. Laser emission from an organic molecular crystal. *Phys. Status Solidi* **13**, 651–655 (1972).
- Hermes, R. E., Allik, T. H., Chandra, S. & Hutchinson, J. A. High-efficiency pyromethene doped solid-state dye lasers. *Appl. Phys. Lett.* **63**, 877–879 (1993).
- Burroughes, J. H. et al. Light-emitting diodes based on conjugated polymers. *Nature* **347**, 539–541 (1990).
- Tang, C. W., VanSlyke, S. A. & Chen, C. H. Electroluminescence of doped organic thin films. *J. Appl. Phys.* **65**, 3610–3615 (1989).
- Hide, F. et al. Semiconducting polymers: a new class of solid-state laser materials. *Science* **273**, 1833–1836 (1996).
- Hide, F., Schwartz, B. J., Diaz-Garcia, M. A. & Heeger, A. J. Laser emission from solutions and films containing semiconducting polymer and titanium dioxide nanocrystals. *Chem. Phys. Lett.* **256**, 424–430 (1996).
- Tessler, N., Denton, D. J. & Friend, R. H. Lasing from conjugated polymer microcavities. *Nature* **382**, 695–697 (1996).
- Kozlov, V. G., Burrows, P. E. & Forrest, S. R. in *Conf. on Lasers and Electro-optics (CLEO '97)* CPD-18 (Opt. Soc. Am., Baltimore, MD, 1997).
- Garbuzov, D. Z., Bulović, V., Burrows, P. E. & Forrest, S. R. Photoluminescence efficiency and absorption of aluminum-tris-quinolate (Alq₃) thin films. *Chem. Phys. Lett.* **249**, 433–437 (1996).
- Pope, M. & Swenberg, C. E. *Electronic Processes in Organic Crystals* (Clarendon, Oxford, 1982).
- Schäfer, F. P. *Dye Lasers* (Springer, Berlin, 1977).
- Coldren, L. A. & Corzine, S. W. *Diode Lasers and Photonic Integrated Circuits* (Wiley, New York, 1995).
- Sokolik, I., Priestly, R., Walser, A. D., Dorsinville, R. & Tang, C. W. Bimolecular reactions of singlet excitons in tris(8-hydroxyquinoline) aluminum. *Appl. Phys. Lett.* **27**, 4168–4170 (1996).
- Brown, A. R. et al. Poly(p-phenylenevinylene) light-emitting diodes: enhanced electroluminescent efficiency through carrier confinement. *Appl. Phys. Lett.* **61**, 2793–2795 (1992).

Acknowledgements. This work was supported by the US Air Force Office of Scientific Research, the National Science Foundation and Universal Display Corporation.

Correspondence and requests for materials should be addressed to V.G.K. (e-mail: kozlov@ee.princeton.edu).

Continuous formation of supported cubic and hexagonal mesoporous films by sol-gel dip-coating

Yunfeng Lu*, Rahul Ganguli*, Celeste A. Drewien*, Mark T. Anderson*, C. Jeffrey Brinker*, Weiliang Gong†, Yongxing Guo†, Hermes Soyez‡, Bruce Dunn‡, Michael H. Huang§ & Jeffrey I. Zink§

* Sandia National Laboratories and The University of New Mexico/NSF Center for Micro-Engineered Materials, The Advanced Materials Laboratory, 1001 University Blvd SE, Albuquerque, New Mexico 87106, USA

† Department of Earth and Planetary Sciences, The University of New Mexico, Albuquerque, New Mexico 87106, USA

Departments of ‡ Materials Science and § Chemistry, University of California, Los Angeles, California 90095-1595, USA

Thin films of surfactant-templated mesoporous materials^{1,2} could find applications in membrane-based separations, selective catalysis and sensors. Above the critical micelle concentration of a bulk silica-surfactant solution, films of mesophases with hexagonally packed one-dimensional channels can be formed at solid-liquid and liquid-vapour interfaces^{3–5}. But this process is slow and the supported films^{3,5} are granular and with the pore channels oriented parallel to the substrate surface, so that transport across the films is not facilitated by the pores. Ogawa^{6,7} has reported a rapid spin-coating procedure for making transparent mesoporous films, but their formation mechanism, microstructure and pore accessibility have not been elucidated. Here we report a sol-gel-based dip-coating method for the rapid synthesis of continuous mesoporous thin films on a solid substrate. The influence of the substrate generates film mesostructures that have no bulk counterparts, such as composites with incipient liquid-crystalline order of the surfactant-silica phase. We are also able to form mesoporous films of the cubic phase, in which the pores are connected in a three-dimensional network that guarantees their accessibility from the film surface. We demonstrate and quantify this accessibility using a surface-acoustic-wave nitrogen-adsorption

technique. We use fluorescence depolarization to monitor the evolution of the mesophase *in situ*, and see a progression through a sequence of lamellar to cubic to hexagonal structures that has not previously been reported.

Continuous films with accessible porosity are frequently cited as a promising application of mesoporous materials^{3–7}. It has been shown^{3–5} that by exceeding the critical micelle concentration, c_{mc} of a bulk silica-surfactant solution, hexagonal liquid crystalline mesophases with one-dimensional pore channels (1-dH) form at solid-liquid and liquid-vapour interfaces by interfacial self-assembly⁸. However, this process is slow and supported films are granular^{3,5}, with an unfavourable orientation of the pore channels parallel to the substrate surface^{3,5}. Ogawa^{6,7} has reported a rapid spin-coating procedure that results in transparent films, but the film-formation mechanism and microstructure have not yet been elucidated. No previous studies have demonstrated accessibility of the mesophase porosity of a supported film.

Our process begins with a homogeneous solution with an initial surfactant concentration $c_0 \ll c_{mc}$. We rely on surfactant enrichment by solvent evaporation to exceed c_{mc} and therefore develop mesophases only during the last few seconds of film deposition (Fig. 1b). By varying c_0 , we have access to a continuous range of silica/surfactant/solvent composition space, enabling us selectively to tailor the film morphology established at the drying line (Fig. 1b).

Precursor solutions were prepared by addition of cationic surfactant (CTAB: $\text{CH}_3(\text{CH}_2)_{15}\text{N}^+(\text{CH}_3)_3\text{Br}^-$) to polymeric silica sols made by a two-step process (A2**)⁹, designed to minimize the siloxane condensation rate¹⁰, promoting facile silica-surfactant co-assembly during film deposition. First, TEOS ($\text{Si}(\text{OC}_2\text{H}_5)_4$), ethanol water and HCl (mole ratios: 1 : 3 : 8 : 5 $\times 10^{-5}$) were refluxed at 60 °C for 90 min. Second, water and HCl were added, increasing the concentration of HCl ([HCl]) to 7.34 mM. After stirring at 25 °C for

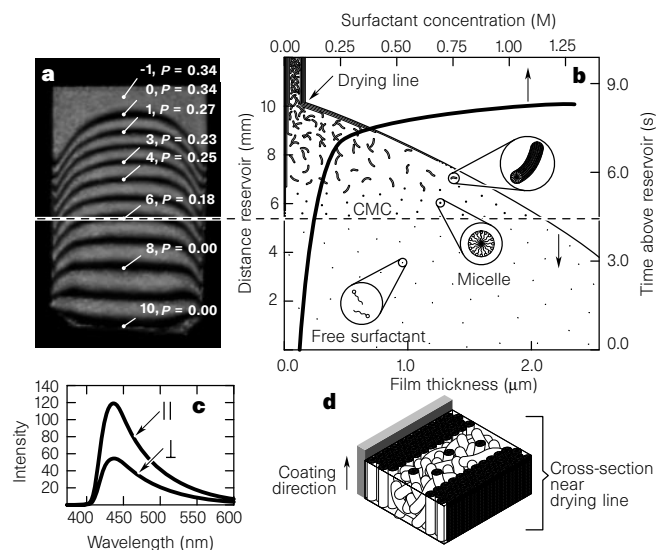
15 min, the sols were aged at 50 °C for 15 min and diluted 1 : 3 with ethanol. Finally, CTAB was added in quantities corresponding to concentrations in the range 0.03–0.11 M (1.5–5.0 wt%). The final reactant mole ratios were: 1 TEOS : 22 $\text{C}_2\text{H}_5\text{OH}$: 5 H_2O : 0.004 HCl : 0.054–0.18 CTAB.

Films were deposited on (100)-silicon by dip-coating at 7.6 cm min^{-1} (Fig. 1). During dip-coating¹¹, the entrained solution experiences preferential evaporation of ethanol¹², progressively enriching the concentrations of water, HCl, and the non-volatile constituents silica and surfactant. After several seconds, the film profile (thickness versus vertical position above the sol reservoir surface) becomes steady in the laboratory frame^{11,12}. Figure 1a shows an interference image of a steady-state film acquired during dip-coating; Fig. 1b shows the corresponding film thickness and surfactant concentration profiles. The surfactant concentration increases by over an order of magnitude, whereas the surfactant/silica ratio remains constant.

Spatially resolved fluorescence-depolarization experiments¹³ using a hydrophobic dye (2,6-TNS) entrained in the coating solution were performed in order to identify micelle formation *in situ* during film deposition (Fig. 1a, c). In the coating solution ($c_0 = 0.06$ M) and within the first ~5 mm above the reservoir surface, the degree of polarization, P , was zero indicating that the probe was freely rotating¹³. At fringe 6, which we identify as c_{mc} , P abruptly increases to 0.18 owing to incorporation of the probe in the hydrophobic micellar interior. Beyond c_{mc} , P continues to increase. In contrast, fluorescence depolarization of films prepared without CTAB gave $P \approx 0$ below the drying line and a finite P appeared only above the drying line. These results show that silica/surfactant micellar structures form only on the substrate by virtue of evaporation-induced surfactant enrichment.

Figure 2a shows X-ray diffraction patterns of supported, as-deposited films prepared with c_0 in the range 0.06–0.11 M; Fig. 2b

Figure 1 a, Optical interference image of the steady-state film, prepared with initial surfactant concentration $c_0 = 0.06$ M during dip-coating at 7.6 cm min^{-1} . The image was acquired using a HeNe laser ($\lambda = 632.8$ nm) and beam expander. Each fringe represents a change in the optical thickness of $\lambda/2(n^2 - \sin^2\theta)^{1/2}$, where n is the refractive index of the depositing sol and θ is the angle of incidence. Superimposed on the interference images are the values of P determined *in situ* as a function of fringe position above the sol reservoir (**c**). Fringes are numbered from 0 at the drying line to 10 near the reservoir surface (position -1 is located in the solidified film about one fringe spacing above the drying line). **b**, Steady-state film cross-section, corresponding to the interference image (**a**), with vertical axes representing distance above the sol reservoir surface and the corresponding time required for the substrate to move that distance. The film-thickness profile was calculated from the interference image assuming a linear refractive index gradient from 1.25 at the liquid meniscus to 1.35 at the drying line. The surfactant concentration was calculated from the thickness profile assuming no surfactant volatility; c_{mc} was established from the fluorescence depolarization experiments described in **a** and **c**. Although the surfactant concentration increases with position/time above the reservoir surface, the surfactant/silica ratio remains fixed. **c**, Fluorescence spectra of 2-*p*-toluidinyl naphthalene-6-sulphonate (2,6-TNS) acquired at fringe position -1. The probe molecule was introduced at an initial concentration of 10^{-4} M and excited at 351 nm with an argon ion laser polarized parallel to the film drawing direction. The beam was focused on a given fringe to a diameter of ~50 μm . Emitted light was analysed with a calcite Glan-Thompson polarizer oriented parallel (pl) or perpendicular (pp) to the incident light. The degree of polarization P was determined from the measured intensities I_{pl} and I_{pp} at 440 nm using the standard relation: $P = (I_{pl} - I_{pp})/(I_{pl} + I_{pp})$. P suddenly increases from 0 to 0.18 at fringe 6 owing to incorporation of the hydrophobic dye into the micellar interior, as verified by control solution experiments of CTAB in ethanol/water mixtures. **d**, Cross-section of an ordered region of the film near the drying line, depicting formation of interfacial regions of liquid crystalline order. 1-dH regions grow from the substrate-film and air-film interfaces towards the film interior. The proposed forma-



tion mechanism involves cooperative assembly of silica-surfactant micellar species with supramolecular surfactant 'templates', for example, cylindrical micelles or hemimicelles, that self-assemble at the substrate-liquid^{3,5} and liquid-air interfaces⁴ at an earlier stage of the deposition process (at or below c_{mc} ; refs 3–5, 8). The interior region shows a disordered worm-like micellar structure¹⁵. Silica completely encases the surfactant assemblies, so that upon surfactant calcination a unimodal pore size is obtained. (Complementary *in situ* AFM experiments performed at the silicon(oxide)-liquid interface using the sol prepared with $c_0 = 0.06$ M (S. Singh, unpublished results) showed the development of an interfacial structure with approximate thickness 0.15 nm over a period of 1 hour, confirming that film growth occurs almost exclusively during dip-coating.)

shows the corresponding patterns after calcination at 400 °C in oxygen. For the films prepared with $c_0 = 0.06$ M, the X-ray diffraction is consistent with a 1-dH mesophase with pore channels oriented parallel to the substrate surface³. X-ray diffraction patterns of the uncalcined films prepared with $c_0 = 0.10$ M and 0.11 M are consistent with an oriented lamellar liquid crystalline mesophase with basal spacing $c = 34.5$ Å (ref. 14). Upon calcination, the film prepared with $c_0 = 0.11$ M collapses to form a substantially amorphous film, whereas the film prepared with $c_0 = 0.10$ M transforms to a new mesostructure comprising both cubic and three-dimensional hexagonal (3-dH) liquid crystalline order. The (111)-peak in Fig. 2b, trace B, unambiguously establishes the cubic phase. The remaining peaks may be indexed on a primitive cubic cell with lattice constant $a = 78.9(12)$ Å or a related 3-dH cell with lattice constants $a = 79.1(5)$ Å and $c = 75.5(6)$ Å.

Figure 3a, b shows cross-sectional transmission electron micrographs (TEM) of calcined films. The film prepared with $c_0 = 0.06$ M exhibits incipient 1-dH order adjacent to the air–film and film–substrate interfaces and a disordered interior region¹⁵ with textural dimensions commensurate with the periodic interfacial structures. By comparison, the film prepared with $c_0 = 0.10$ M exhibits a uniform silica mesophase characterized by a diamond-shaped

lattice unique to the 3-dH hexagonal mesophase. Representative plan-view images are shown in Fig. 3c, d and 4c. The incipient 1-dH film is composed of swirling patterns of organized tubules⁵ separated by disordered regions similar to the film interior. Swirling is consistent with the low bending energy of the tubules⁵ and the inability of the liquid–vapour interface to impose long-range order on the tubule assembly process. In contrast, the cubic 3-dH film has extensive regions of three-dimensional ordering (image acquired through complete film thickness), with seamless transitions between ordered ‘domains’.

To avoid any effects due to constrained one-dimensional shrinkage of the film normal to the substrate direction during calcination³, film fragments were detached from the substrate before heating. No structures unique to the hexagonal phase have been observed in detached fragments. The TEM of a detached fragment (Fig. 4a) shows a square periodic lattice with spacing 79 Å, consistent with the (100)-orientation of the primitive cubic phase; Fig. 4b shows a simulated image with $a = 79$ Å.

We explain the formation of incipient 1-dH films as follows (Fig. 1b, d). Beginning with a homogeneous solution ($c_0 \ll c_{mc}$), preferential ethanol evaporation enriches the depositing film in silica, surfactant and water, causing c to exceed c_{mc} at a CTAB

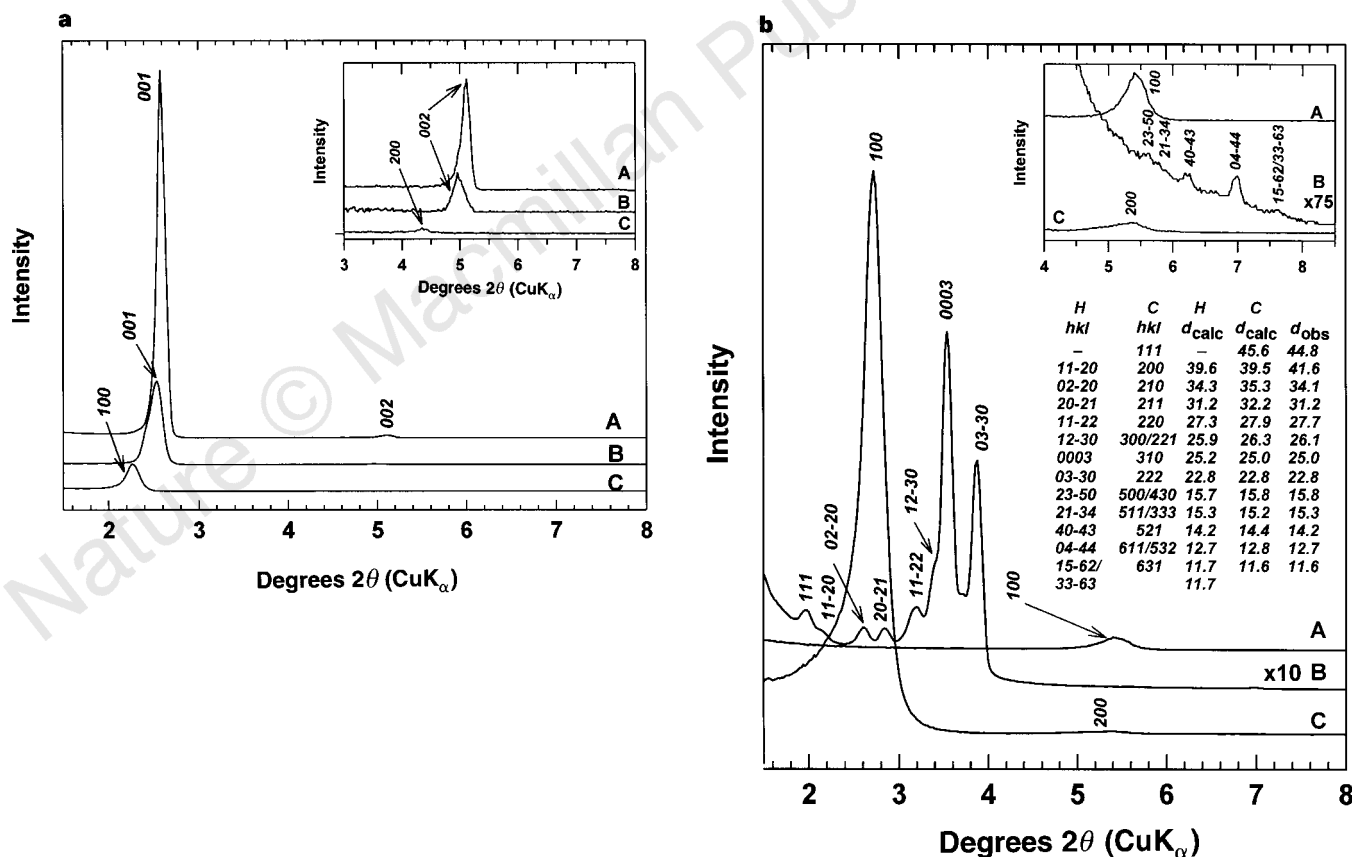


Figure 2 a, X-ray diffraction spectra (A–C) recorded on a Siemens D500 diffractometer using Ni-filtered $\text{CuK}\alpha$ radiation with $\lambda = 1.5418$ Å in θ – 2θ scan mode for uncalcined silica/surfactant mesophases prepared with A, $c_0 = 0.11$ M (5.0 wt% CTAB); B, $c_0 = 0.10$ M (4.2 wt% CTAB); and C, $c_0 = 0.06$ M (2.5 wt% CTAB). Inset: region where 2θ is 2.5–10°, with an intensity scale expanded 115 times, shows that all three spectra contain second-order reflections at $d = 1/2d_{\text{primary}}$. b, XRD patterns of samples in a after calcination at 400 °C (heating rate of 1 °C min^{−1}). (Inset: region where 2θ is 3.5–10°, with an intensity scale expanded 15 times). Trace A (see a) is consistent with a lamellar liquid crystalline mesophase (L) with basal spacing $c = 34.5$ Å. The L mesophase collapses upon calcination (this panel), leaving only a tiny residual peak (0.5% of the intensity of the (001) peak of the uncalcined film) indicative of a small fraction of a non-collapsed, porous phase with $d \approx 16$ Å. Trace B (see a) resembles the L meso-

phase of trace A, but upon calcination, the L film transforms to a new mesostructure composed of primitive cubic and 3-d hexagonal (3-dH) mesophases (this panel). (Apart from the (111) reflection, the peaks in trace B are indexed according to the 3-dH mesophase; the corresponding cubic indices (C) are shown in the table in the figure.) The cubic mesophase is characterized by $a = 78.9(12)$ Å, and the 3-dH mesophase by $a = 79.1(5)$ Å and $c = 75.5(6)$ Å. Note basal (0003) and prismatic (03-30) or cubic (222) preferred orientation of the C/3-dH film. Also the (111) peak is not indexable on the 3-dH unit cell, providing unique evidence for the cubic mesophase. Trace C is indicative of a 1-dH mesophase with a unit cell constant $a = 45.8$ Å (uncalcined) and $a = 37.8$ Å (calcined), in which the axes of the pore channels are oriented parallel to the substrate surface (note the absence of mixed reflections (110) and (210)³).

concentration of ~ 0.1 M. Beyond c_{mc} , incipient liquid-crystalline domains 'grow' from both interfaces towards the centre (with an attendant increase in P) by co-assembly of silica-surfactant micellar species with the preorganized solid-liquid^{3,5,16} and liquid-vapour interfacial regions⁴. Solidification at the drying line establishes a structure composed of disordered cylindrical micelles¹⁵ sandwiched between ordered regions of hexagonal mesophases.

With increasing c_0 (but still below c_{mc}), we observe incipient cubic (R.G., unpublished results) and lamellar mesophases that are consistent with a progressive evolution of the structure in response

to surfactant enrichment. For $c_0 \geq 0.10$ M, we believe homogeneous lamellar films (Fig. 4d) are nucleated by interfacially organized surfactant bilayers¹⁶, observed using atomic force microscopy (AFM) for 0.11 M CTAB sols equilibrated with (100)-silicon (S. Singh, unpublished results). We suggest that cubic 3-dH films form by a lamellar $\rightarrow$ cubic $\rightarrow$ hexagonal pathway driven by continuing condensation of the siloxane framework as proposed¹⁷ to explain lamellar to hexagonal transformations¹⁸ in bulk silica mesophases. Although unreported for inorganic mesophases, lamellar to cubic transformations are common in CTAB¹⁴ and

Figure 3 Cross-sectional TEM images (recorded with the silicon substrate oriented along the [110] zone axis) of: **a**, a calcined incipient 1-dH film, the same film used to obtain trace C in Fig. 2b. Periodic ordered regions (with different thicknesses) are observed at both the substrate-film and film-vapour (glue) interfaces mainly as stripes that transform into 'brick-like' features³ for those domains with channel axes oriented approximately normal to the sectioning direction. Brick-like structures presumably result from constrained one-dimensional shrinkage of the films upon calcination³. **b**, the cubic 3-dH film prepared with $c_0 = 0.10$ M, corresponding to trace B in Fig. 2b. The film is substantially ordered throughout its thickness. Electron diffraction and simulation uniquely identify this structure as the $[\bar{1}123]$ -zone direction of the 3-dimensional hexagonal mesophase. Note the observed $68 \pm 2^\circ$ angle between the (1010) and (0111) reflections. The calculated angle is 68.3° . This image cannot be accounted for by a $78.9\text{-}\text{\AA}$ cubic cell. **c**, Plan-view image corresponding to the film-vapour interface for the same sample, as in **a** (recorded under identical conditions but with silicon substrate oriented along the [001] zone axis). The swirling patterns of tubule bundles show no preferred orientation and are similar in appearance to those observed by Aksay *et al.*⁵ for silica-surfactant films grown at the silica-water interface. Interfaces between the swirls appear disordered but exhibit textural features comparable in dimension to the periodically ordered domains with no apparent gaps. **d**, Plan-view of cubic 3d-H film imaged through the complete film thickness and a thin layer of the underlying silicon substrate (note bend contours), showing seamless transition at a bend contour consistent with a continuous cubic 3-dH film. **e**, Scanning electron micrograph image of the incipient HLC film surface showing absence of granularity on macroscopic scales. (TEM samples were prepared by conventional sectioning techniques comprising grinding and polishing, followed by dimpling, and ion milling until electron-transparent.)

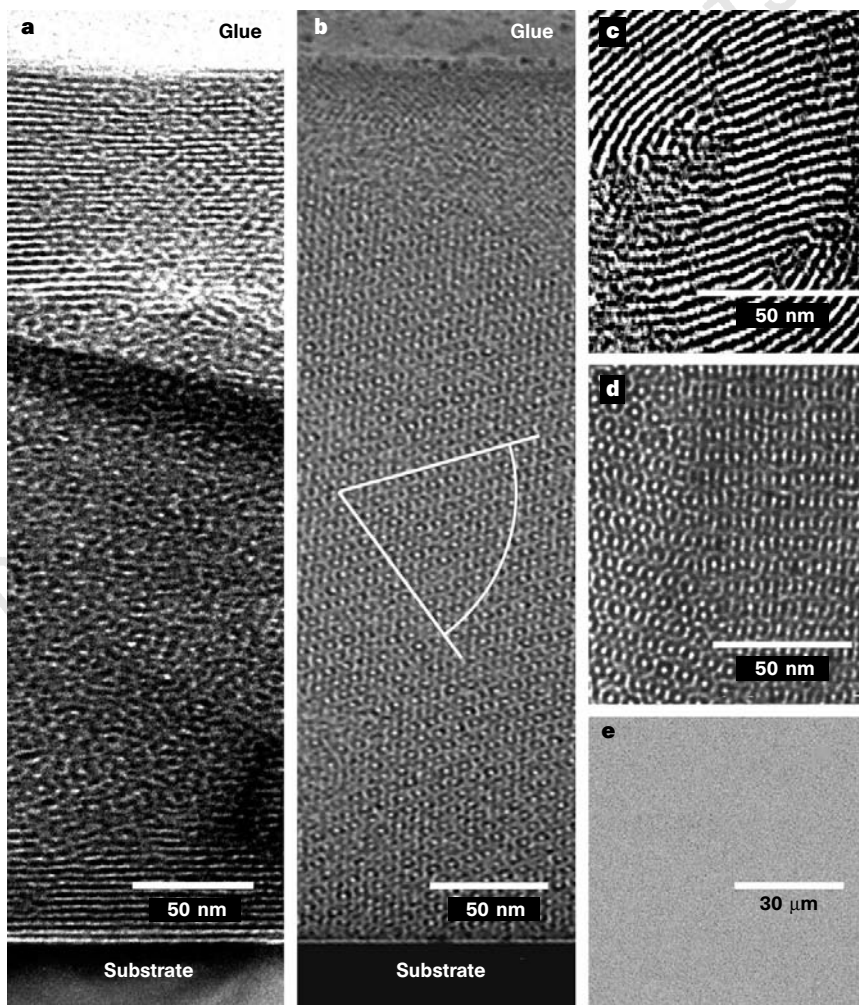
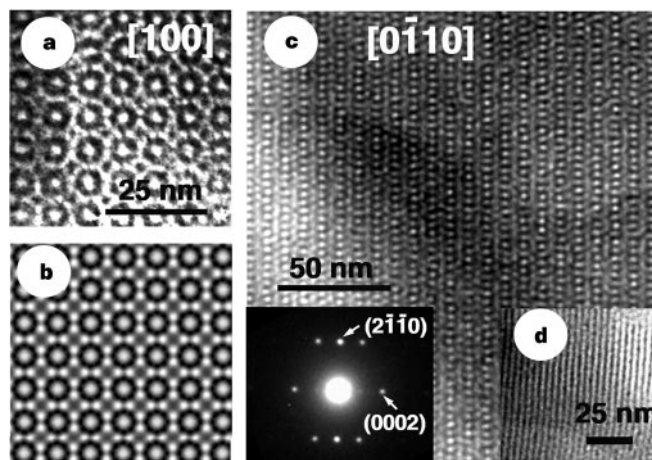


Figure 4 **a**, TEM image of the [100]-zone of a calcined film fragment prepared with $c_0 = 0.10$ M and detached from the substrate before heating. The image shows the orthogonal arrangement of pores, with a pore-to-pore distance of $\sim 79\text{\AA}$, consistent with a cubic lattice. **b**, Simulated image of the [100]-oriented cubic mesophase, showing approximately the same phase contrast as the actual image. **c**, Plan-view of cubic 3d-H film imaged through the film and silicon substrate (not bend contours), showing a large region of [01-10] orientation. Inset: selected area defraction pattern of the [01-10] zone axis. Note this image and the selected area diffraction pattern are also consistent with the [210]-cubic orientation. **d**, Layered lamellar structure of the film prepared with $c_0 = 0.10$ M before transformation to the cubic phase.



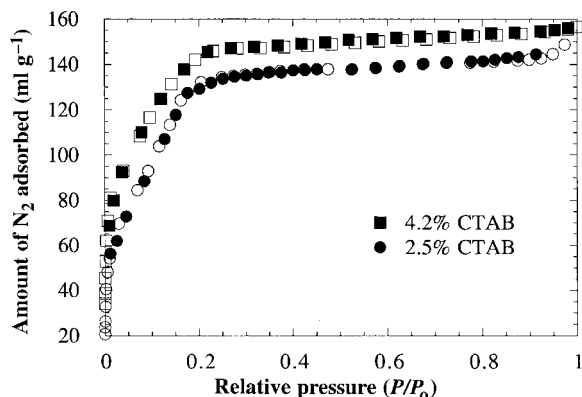


Figure 5 N_2 sorption isotherms obtained at 77 K for a 400-nm thick incipient 1-dH film prepared with $c_0 = 0.06$ M (adsorption, white circles; desorption, black circles) and a 250-nm-thick cubic 3-dH film prepared with $c_0 = 0.10$ M (adsorption, white squares; desorption, black squares). The films were applied to ~ 1 cm² area of a piezoelectric ST-cut quartz substrate with interdigital gold transducers designed to operate at ~ 97 MHz. Mass change was monitored (~ 80 pg cm⁻² sensitivity) as a function of relative pressure using a surface acoustic wave technique²¹.

membrane lipid/water systems¹⁹. It has been suggested that two lipid bilayers transform through an inverted micellar intermediate to a cubic mesophase²⁰. We propose that the 3-dH film is derived from the parent cubic film by constrained one-dimensional shrinkage during calcination². The cubic $\rightarrow$ 3-dH transformation can be considered as a constrained one-dimensional distortion of the cubic cell along the [110]-direction. Comparable lattice parameters between the two mesophases allow seamless transitions between ordered domains and cause cubic and 3-dH domains to be essentially indistinguishable in plan-view (Fig. 4c).

A surface acoustic wave technique²¹ was used to determine pore accessibility of supported films. Figure 5 compares nitrogen sorption isotherms of the incipient 1-dH and cubic 3-dH films. Despite the substantially different degrees of ordering, the isotherms are qualitatively similar. The lack of hysteresis and absence of any appreciable adsorption at relative pressures above 0.3 is consistent with a unimodal porosity with no interparticle meso- or macroporosity²². The surface areas calculated for the cubic 3-dH and incipient 1-dH films are 734 and 648 m² g⁻¹, respectively, demonstrating the accessibility of the mesophase porosity. In addition, the trans-film flux of cubic 3-dH films prepared as supported membranes increased by over 1,000 $\times$ upon calcination, establishing through-thickness pore connectivity (A. Tsai, unpublished results).

We have demonstrated a rapid, continuous process, enabling the practical utilization of mesostructures in thin-film form. The uniform three-dimensional pore channel systems of cubic 3-dH films and the absence of granularity suggest applications in molecular separation, catalysis and sensors. The dip-coating procedure combined with the optical probe technique enables us to follow the progressive evolution of the mesostructured films and should provide insight into the synthesis of complex, self-organized organic/inorganic assemblies in general²³. □

Received 24 June; accepted 16 July 1997.

- Kresge, C. T., Leonowicz, M. E., Roth, W. J., Vartuli, J. C. & Beck, J. S. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
- Beck, J. S. *et al.* A new family of mesoporous molecular sieves prepared with liquid crystal templates. *J. Am. Chem. Soc.* **114**, 10834–10843 (1992).
- Yang, H., Kuperman, A., Coombs, N., Mamiche-Afara, S. & Ozin, G. A. Synthesis of oriented films of mesoporous silica on mica. *Nature* **379**, 703–705 (1996).
- Yang, H., Coombs, N., Sokolov, I. & Ozin, G. A. Free-standing and oriented mesoporous silica films grown at the air–water interface. *Nature* **381**, 589–592 (1996).
- Aksay, I. A. *et al.* Biomimetic pathways for assembling inorganic thin films. *Science* **273**, 892–898 (1996).
- Ogawa, M. Formation of novel oriented transparent films of layered silica-surfactant nanocomposites. *J. Am. Chem. Soc.* **116**, 7941–7942 (1994).

- Ogawa, M. A simple sol-gel route for the preparation of silica-surfactant mesostructured materials. *Chem. Commun.* 1149–1150 (1996).
- Manne, S. & Gaub, H. E. Molecular organization of surfactants at solid–liquid interfaces. *Science* **270**, 1480–1482 (1995).
- Brinker, C. J. *et al.* in *Access in Nanoporous Materials* (eds Pinnavaia, T. J. & Thorpe, M. F.) 123–139 (Plenum, New York, 1995).
- Brinker, C. J. & Scherer, G. W. in *Sol–Gel Science* 120 (Academic, San Diego, 1990).
- Manne, S., Hurd, A. J., Schunk, P. R., Frye, G. C. & Ashley, C. S. Review of sol-gel thin film formation. *J. Non-Cryst. Solids* **147**, 424–436 (1992).
- Nishida, F. *et al.* In situ fluorescence probing of the chemical-changes during sol-gel thin-film formation. *J. Am. Ceram. Soc.* **78**, 1640–1648 (1995).
- Shinitzky, M., Dianoux, A. C., Gitler, C. & Weber, G. Microviscosity and order in the hydrocarbon region of micelles and membranes determined with fluorescent probes: I. Synthetic micelles. *Biochemistry* **10**, 2106–2113 (1971).
- Auvray, X., Petipas, C., Anthore, R., Rico, I. & Lattes, A. X-ray diffraction study of mesophases of cetyltrimethylammonium bromide in water, formamide, and glycerol. *J. Phys. Chem.* **93**, 7458–7464 (1989).
- Ryoo, R., Kim, J. M., Ko, C. H. & Shin, C. H. Disordered molecular sieve with branched mesoporous channel network. *J. Phys. Chem.* **100**, 17718–17721 (1996).
- Pashley, R. M. & Israelachvili, J. N. A comparison of surface forces and interfacial properties of mica in purified surfactant solutions. *Colloids Surf.* **2**, 169–187 (1981).
- Monnier, A. *et al.* Cooperative formation of inorganic organic interfaces in the synthesis of silicate mesostructures. *Science* **261**, 1299–1303 (1993).
- Israelachvili, J. N., Mitchell, D. J. & Ninham, B. W. Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers. *J. Chem. Soc.* **2**, 1525–1568 (1976).
- Lindblom, G. & Reifors, L. Cubic phases and isotropic structures formed by membrane lipid—possible biological relevance. *Biochim. Biophys. Acta* **988**, 221–256 (1989).
- Siegel, D. Inverted micellar structures in bilayer membranes. *J. Biophys.* **45**, 399–420 (1986).
- Frye, G. C., Riccio, A. J., Martin, S. J. & Brinker, C. J. Characterization of the surface area and porosity of sol-gel films using SAW devices. *Mat. Res. Soc. Symp. Proc.* **121**, 349–354 (1988).
- Greg, S. J. & Sing, K. S. W. *Adsorption, Surface Area, and Porosity* 2nd edn (Academic, New York, 1982).
- Mann, S. & Ozin, G. A. Synthesis of inorganic materials with complex form. *Nature* **382**, 313–318 (1996).

Acknowledgements. We thank S. Singh, A. Tsai, M. Rodriguez, M. Eatough and R. Tissot for assistance with experiments and M. Aragon for technical illustrations. This work was partially supported by the UNM/NSF Center for Micro-Engineered Materials, by grants from the NSF, by the DOE Basic Energy Sciences, the Electric Power Research Institute and the DOE Federal Energy Technology Center. This work was done under contract from the US Department of Energy. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the US Department of Energy.

Correspondence and requests for materials should be addressed to C.J.B. (e-mail: cjbrink@sandia.gov).

Extraction of a hydrophilic compound from water into liquid CO₂ using dendritic surfactants

A. I. Cooper*, J. D. Londono†‡, G. Wignall†, J. B. McClain*, E. T. Samulski*, J. S. Lin†, A. Dobrynin*, M. Rubinstein*, A. L. C. Burke*, J. M. J. Fréchet§ & J. M. DeSimone*

* Department of Chemistry, CB#3290, Venable and Kenan Laboratories, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290, USA

† Oak Ridge National Laboratory, Oak Ridge, Tennessee 37381-6031, USA

‡ Department of Chemical Engineering, University of Tennessee, Knoxville, Tennessee 37996-220, USA

§ Department of Chemistry, University of California-Berkeley, Berkeley, California 94720, USA

Dendrimers are well defined, highly branched polymers^{1–5} that adopt a roughly spherical, globular shape in solution. Their cores are relatively loosely packed and can trap guest molecules^{5–7}, and by appropriate functionalization of the branch tips the macromolecules can act as unimolecular micelle-like entities⁶. Here we show that dendrimers with a fluorinated shell are soluble in liquid carbon dioxide and can transport CO₂-insoluble molecules into this solvent within their cores. Specifically, we demonstrate the extraction of a polar ionic dye, methyl orange, from water into CO₂ using these fluorinated dendrimers. This observation suggests possible uses of such macromolecules for the remediation of contaminated water, the extraction of pharmaceutical products from fermentation vessels, the selective encapsulation of drugs for targeted delivery^{6,7} and the transport of reagents for chemical

reactions (such as polymerization^{8–11}) in liquid and supercritical CO₂ solvents.

Although CO₂ is a good solvent for many small molecules, only two classes of polymers have shown significant solubility (>10%) in CO₂ under practicable conditions (<100 °C and <350 bar); amorphous (and low-melting) fluoropolymers^{8,12} and polysiloxanes^{8,13}. As a result, there has been considerable effort to design fluorinated

and siloxane-based surfactants that can stabilize dispersions of otherwise insoluble polymeric materials in carbon dioxide^{9,10,13,14}. Small-angle X-ray scattering (SAXS) and small-angle neutron scattering (SANS) studies have shown that partially fluorinated amphiphilic surfactants can aggregate into micelles in carbon dioxide solution, the precise number of polymer chains per micelle being influenced by the CO₂ density^{15–17}.

The micelle-like structure investigated here is quite different. By extending methods developed by Meijer⁶, a fourth-generation hydrophilic dendrimer¹⁸, DAB-dendr-(NH₂)₃₂, **1**, was functionalized with a 'CO₂-philic' shell, derived from a heptamer acid fluoride of hexafluoropropylene oxide **2** CF₃CF₂CF₂(OCF(CF₃)CF₂)₅-OCF(CF₃)C(O)F, thus generating what may be considered as a well-defined, CO₂-soluble, unimolecular dendritic micelle, **3** (compounds **1**, **2** and **3** are shown in Fig. 1). Characterization data agreed with a structure where, on average, 90% of the peripheral amine groups of **1** had been functionalized with perfluoropolyether chains. NMR spectroscopy showed that the NHCO proton signal for **3** (8.70 p.p.m.) was shifted significantly downfield from the corresponding signal at 7.75 p.p.m. for the single-chain model compound, F₃CF₂CF₂O(CF(CF₃)CF₂O)₅(CF(CF₃)C(O)NH-CH₂CH₂CH₃), **4**. This is consistent with strong intramolecular hydrogen bonding between amide groups in the closely packed dendritic structure of **3**, as described previously for analogous systems⁶. We found the functionalized dendrimer **3** to be insoluble in water (<10 p.p.m.) and most common organic solvents (methanol, methylene chloride, tetrahydrofuran, hexanes, chloroform and acetone), but to be soluble at room temperature in liquid CO₂ at pressures above 76 atm (1,110 pounds per square inch). Dendrimer **1** was insoluble in CO₂ under the same conditions.

As many highly polar molecules are insoluble in carbon dioxide, we decided to investigate the potential of **3** as a host system for polar guests in CO₂. It has been demonstrated previously that guest molecules can be trapped in a permeable dendritic core *in situ* by constructing a dense shell around the impregnated core and generating a 'dendritic box'^{19–21}. Alternatively, the dendritic core of a pre-formed, micelle-like host can be loaded with guest molecules in a homogeneous solution of both host and guest⁶. By contrast, we have concentrated on the micelle-assisted transfer of guest molecules between two immiscible phases. Quite surprisingly, we have

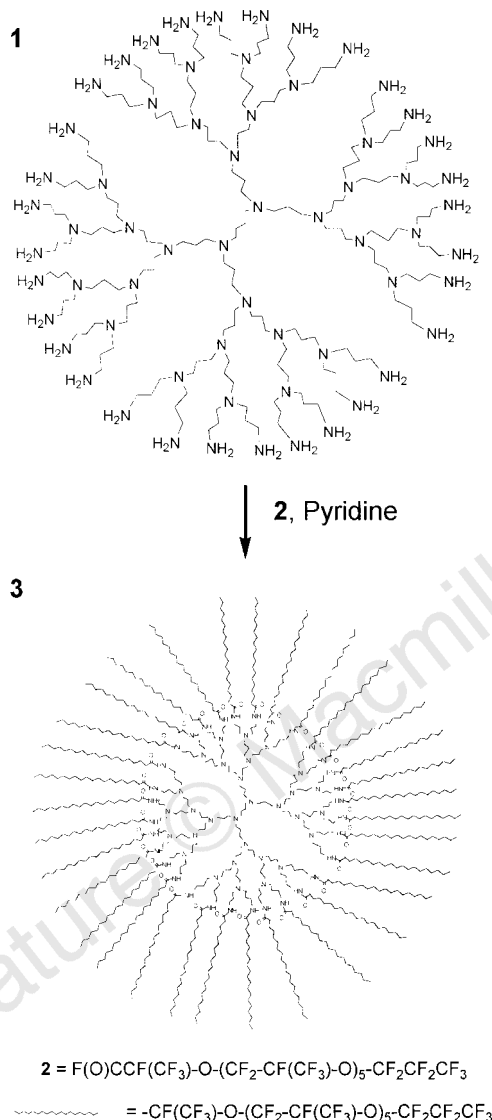


Figure 1 Synthesis of the unimolecular dendritic micelle. A solution of 2.92 g of **2** (35.2 equiv., 2.28 mmol) in freon 113 (50 ml) was added under argon to 0.25 g of DAB-dendr-(NH₂)₃₂ **1** (0.071 mmol) and 0.20 g pyridine (35.2 equiv., 2.28 mmol). The dendrimer **1** was insoluble in freon but was quite rapidly solubilized as the substitution reaction proceeded heterogeneously. Isolated yield after purification, 93%. ¹H NMR (freon 113/acetone-*d*₆): 8.70 (br, s, CONH-CH₂), 3.45 (br, s, CONH-CH₂), 2.40 (br, s, CH₂-N), 1.75 (br, s, NCH₂CH₂CH₂N + NCH₂CH₂CH₂CH₂N). ¹⁹F NMR (freon 113/acetone-*d*₆): -76.72 to 79.81 (m, (O-CF(CF₃)CF₃)), -127.06, -127.48 (CF₃CF₂CF₂O), -129.21 (m, CF(CF₃)C(O)NHCH₂), ±142.32 (br, m, (O-CF(CF₃)CF₂)₅). Infrared (film): amide N-H stretch 3,350.7 cm⁻¹; C-H sat. 2,956.6, 2,835.9 cm⁻¹; amide C=O 1,709.5 cm⁻¹; N-H bend 1,540.9 cm⁻¹; C-F stretch 1,307.6, 1,235.2, 1,203.1, 1,146.8 cm⁻¹; C-O stretch 985.9 cm⁻¹. Ultraviolet-visible (freon 113): wavelengths 240, 273 nm. Differential scanning calorimetry: phase transitions at -58.5 °C and +27.6 °C. Elemental microanalysis: calculated for C₇₉₈H₄₀₃F₁₁₈₁N₆₂O₂₀₂ (average 90% substitution) C, 26.01%; H, 1.11%; N, 2.38%; F, 61.63%; found C, 25.24%; H, 1.17%; N, 2.68%; F, 59.24%. Relative molecular mass (*M*_r): estimated from elemental analysis/¹H NMR (90% substitution), 36.4 (± 1.0) × 10³; determined by small angle X-ray scattering, 33.5 (± 3.0) × 10³.

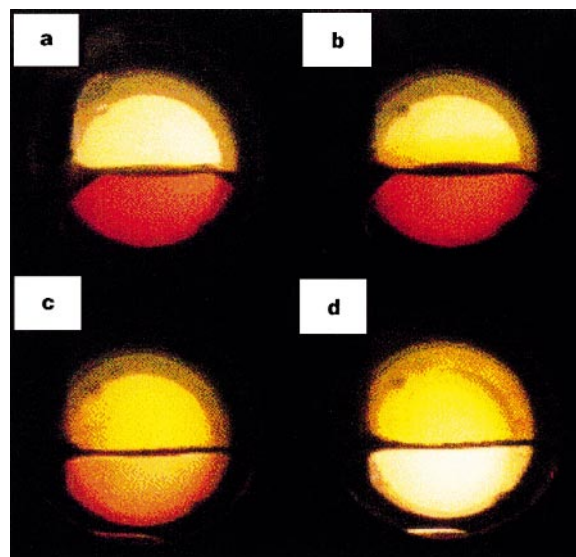


Figure 2 Diffusion of methyl orange (**5**) from aqueous solution (upper phase in figure) into a unimolecular dendritic micelle **3** in liquid CO₂ (lower phase) at 23.5 °C, 340 atm. Video images taken 1 min (**a**), 6 min (**b**), 30 min (**c**) and 150 min (**d**) after the addition of CO₂. Initial molar ratio of **5** : **3** = 1 : 60.

found that the dendritic micelle **3** can transfer methyl orange (**5**), a CO₂-insoluble ionic dye²², from aqueous solution into carbon dioxide (Fig. 2). A small quantity of **3** was weighed onto one of the CaF₂ windows of a 2.5-ml high-pressure cell, the cell assembled, and an aqueous solution of **5** ($3.30 \times 10^{-3} \text{ mol l}^{-1}$) was added such that it did not come into contact with **3**. On addition of CO₂, the dendritic micelle dissolved rapidly, forming a colourless CO₂ phase (Fig. 2a). With time and in the absence of agitation, the colour in the aqueous phase decreased in intensity and a gradual colouring of the CO₂ layer was observed (Fig. 2b, c). After 150 min, it appeared that no dye remained in the aqueous phase (Fig. 2d). When the cell was depressurized, the dye-loaded micelle was deposited on the interior of the cell as a yellow film.

Spectroscopic analysis detected no dye in the residual aqueous phase, which suggests that under these conditions (a large excess of **3**) it was possible to extract this dye completely from the aqueous layer. No coloration of the CO₂ phase was observed in the absence of **3**. The wavelength of the ultraviolet absorption for **5** in the dendritic core of **3**/CO₂ (Fig. 3a) was found to be significantly shifted from the absorption of **5** in aqueous solution (425–430 nm compared to 464 nm), and also from the wavelength observed for **5** dissolved in a water-in-CO₂ microemulsions^{23,24}. This wavelength shift is also evident from the difference in colour of the aqueous and CO₂ phases in Fig. 2. To calculate the maximum number of dye molecules that could occupy the dendritic core, experiments were conducted with varying molar ratios of **5** and **3** in the aqueous/CO₂ phases (Fig. 3b). In all cases, the growth of the ultraviolet absorption of **5** was monitored as a function of time until the absorbance reached a constant value, usually after 25–35 h at room temperature without stirring. The maximum integral absorbance was related to the concentration of the dye in the CO₂ phase, which was then used to calculate the number of dye molecules per micelle core. At low ratios of dye to micelle (for example, 1 : 1), it was possible to extract essentially all of the dye from the aqueous layer and, indeed, an average value of around one dye molecule per dendritic core was determined spectroscopically. This is important because it suggests that the extinction coefficient of the dye absorption does not change

significantly in the dendrimer core, and also that our quantification methods are valid⁶. As the molar ratio of **5** to **3** was increased, the average dye loading increased accordingly, up to a maximum of around 12 dye molecules per dendritic core. Equivalent experiments were conducted with the dye rose bengal, **6** (Fig. 3b). The maximum number of dye molecules per dendritic core (around seven) was consistent with the somewhat larger size of this guest molecule (relative molecular mass, *M_r*, of **5**, 327.37; *M_r* of **6**, 1017.65), and was also comparable with Meijer's observations for alkyl modified dendrimers⁶.

The precise mechanism by which the dyes migrate from the aqueous phase into the dendritic core in CO₂ is not yet known. As dyes **5** and **6** are essentially insoluble in the CO₂ phase and **3** is essentially insoluble in water, it seems conceivable that the highly plasticized dendritic micelle might adopt a distorted conformation at the water–CO₂ interface, such that the hydrophilic core comes in close proximity to the aqueous layer. Preliminary experiments using Reichardt's dye²⁵ as a solvatochromic probe for the effective polarity of the dendritic micelle interior²⁶ suggest that the diffusion of significant amounts of water into the micelle core may accompany the phase transfer of both **5** and **6**.

Small angle X-ray scattering measurements^{16,27}, using a high-pressure cell similar to that described previously²⁸, were used to estimate the radius of gyration and the relative molecular mass of the fluorinated dendrimer. Standard corrections were applied and the data were normalized to units of absolute differential scattering cross-section²⁹, $d\Sigma/d\Omega$ (Q) (cm⁻¹) where Σ denotes the total scattering cross-section, $d\Omega$ is the solid angle and Q is the scattering vector. Scattering data were collected at 340 atm and 25 °C for a number of concentrations of **3** in CO₂. Data were fitted to a variable arm number star model (an 'f-arm star model') over the whole range of Q, and linear fits were also obtained for $[d\Sigma/d\Omega(Q)]^{-1}$ versus Q^2 , in the range $0.01 \leq Q \leq 0.05 \text{ \AA}^{-1}$. The radius of gyration, *R_g*, derived from the f-arm star fits for **3** in CO₂ was $30.0 \pm 1.0 \text{ \AA}$. The value of *R_g* did not vary significantly with concentration, which suggests that no particle aggregation occurred in the CO₂ solution under these conditions. A plot of

Figure 3 a, Ultraviolet–visible spectra illustrating the diffusion of methyl orange (**5**) from an aqueous phase into a unimolecular dendritic micelle **3** in liquid CO₂ (24.9 °C, 340 atm). Spectra taken every 30 min. The high-pressure cell was arranged so that the ultraviolet beam passed only through the CO₂ phase. Molar ratio of **5** : **3** = 60 : 1. **b**, Maximum number of guest molecules per micelle core (*n_{max}*) for methyl orange (trace labelled '5') and rose bengal (trace labelled '6') versus the initial molar ratio (*n_{init}*) of **5** : **6**.

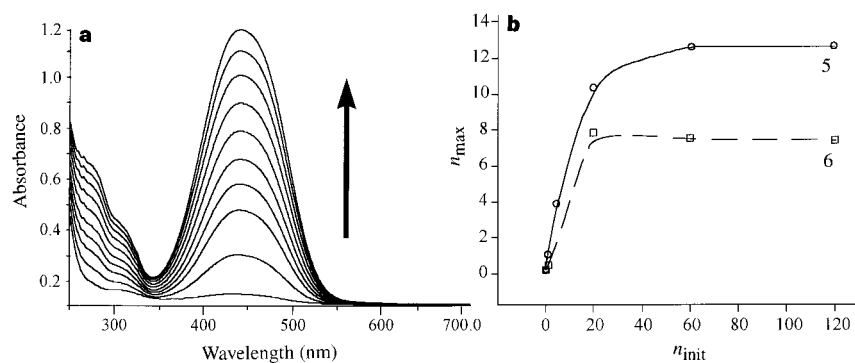
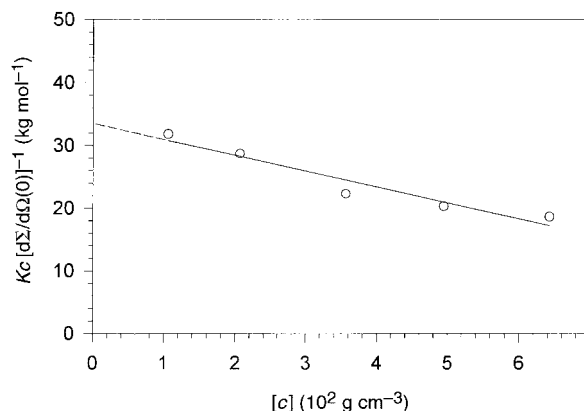


Figure 4 Plot of $K_c[d\Sigma/d\Omega(0)]^{-1}$ versus concentration (*c*) for the dendritic micelle **3** in liquid CO₂ (23.7 °C, 340 atm); see text. $K = (\rho_p - \rho_s)^2 A_0 V_p c / [d\Sigma/d\Omega(0)]$, where ρ_p and ρ_s are the scattering length densities of particle and solvent, *A₀* is Avogadro's number and *V_p* is the estimated particle volume.



$Kc[d\Sigma/d\Omega(0)]^{-1}$ versus concentration of **3** is shown in Fig. 4, based on the $d\Sigma/d\Omega(0)$ values derived from f-arm star fits. (Here K is the SAXS contrast factor and c is the concentration.) The intercept at zero concentration gives the weight average relative molecular mass for **3** to be $33.5(\pm 3.0) \times 10^3$ which is consistent with values obtained by ^1H NMR and elemental microanalysis ($36.4(\pm 1.0) \times 10^3$). The M_r value derived from linear (Zimm) fits was within 0.3% of the value derived from the f-arm star fits.

These experiments demonstrate that surfactant-modified CO_2 can be used to enhance dramatically the applicability of supercritical-fluid extraction applications using environmentally benign CO_2 to replace some of the billions of pounds of organic solvents used every year in extraction and cleaning applications. $\square$

Received 4 April; accepted 7 July 1997.

- Tomalia, D. A. *et al.* A new class of polymers: starburst-dendritic macromolecules. *Polym. J.* **17**, 117–132 (1985).
- Hawker, C. J., Lee, R. & Frechet, J. M. J. One-step synthesis of hyperbranched dendritic polyesters. *J. Am. Chem. Soc.* **113**, 4583–4588 (1991).
- Frechet, J. M. J. Functional polymers and dendrimers: reactivity, molecular architecture, and interfacial energy. *Science* **263**, 1710–1715 (1994).
- Knapen, J. W. J. *et al.* Homogeneous catalysts based on silane dendrimers functionalized with arylnickel(II) complexes. *Nature* **372**, 659–663 (1994).
- Newkome, G. R., Moorefield, C. N. & Vogtle, F. *Dendritic Molecules: Concepts, Syntheses, Perspectives* (VCH, Weinheim, 1996).
- Stevelmans, S. *et al.* Synthesis, characterization and guest-host properties of inverted unimolecular dendritic micelles. *J. Am. Chem. Soc.* **118**, 7398–7399 (1996).
- Hawker, C. J., Wooley, K. L. & Frechet, J. M. J. Unimolecular micelles and globular amphiphiles: dendritic macromolecules as novel recyclable solubilization agents. *J. Chem. Soc. Perkin I* 1287–1297 (1993).
- DeSimone, J. M., Guan, Z. & Eisbernd, C. S. Synthesis of fluoropolymers in supercritical carbon dioxide. *Science* **257**, 945–947 (1992).
- DeSimone, J. M. *et al.* Dispersion polymerizations in supercritical carbon dioxide. *Science* **265**, 356–359 (1994).
- Hsiao, Y.-L., Maury, E. E., DeSimone, J. M., Mawson, S. & Johnston, K. P. Dispersion polymerization of methyl methacrylate stabilized with poly(1,1-dihydroperfluorooctylacrylate) in supercritical carbon dioxide. *Macromolecules* **28**, 8159–8166 (1995).
- Cooper, A. I. & DeSimone, J. M. Polymer synthesis and characterization in liquid/supercritical carbon dioxide. *Curr. Opin. Solid State Mater. Sci.* **1**, 761–768 (1996).
- McHugh, M. A. & Krukonis, V. J. *Supercritical Fluid Extraction* 2nd edn (Butterworth-Heinemann, Stoneham, MA, 1994).
- Shaffer, K. A., Jones, T. A., Canelas, D. A. & DeSimone, J. M. Dispersion polymerizations in carbon dioxide using siloxane-based stabilizers. *Macromolecules* **29**, 2704–2706 (1996).
- Canelas, D. A., Betts, D. E. & DeSimone, J. M. Dispersion polymerization of styrene in supercritical carbon dioxide: importance of effective surfactants. *Macromolecules* **29**, 2818–2821 (1996).
- McClain, J. B. *et al.* Design of nonionic surfactants for supercritical carbon dioxide. *Science* **274**, 2049–2052 (1996).
- Fulton, J. L. *et al.* Aggregation of amphiphilic molecules in supercritical carbon dioxide: a small angle X-ray scattering study. *Langmuir* **11**, 4241–4249 (1995).
- McClain, J. B. *et al.* Solution properties of a CO_2 -soluble fluoropolymer via small angle neutron scattering. *J. Am. Chem. Soc.* **118**, 917–918 (1996).
- de Brabander-van den Berg, E. M. M. & Meijer, E. W. Poly(propylene imine) dendrimers: large-scale synthesis by heterogeneously catalyzed hydrogenations. *Angew. Chem. Int. Edn Eng.* **32**, 1308–1311 (1993).
- Jansen, J. F. G. A., de Brabander-van den Berg, E. E. M. & Meijer, E. W. Encapsulation of guest molecules into a dendritic box. *Science* **266**, 1226–1229 (1994).
- Bosman, A. W., Jansen, J. F. G. A., Jansen, R. A. J. & Meijer, E. W. Charge transfer in the dendritic box. *Polym. Mater. Sci. Eng.* **73**, 340–341 (1995).
- Jansen, J. F. G. A., Meijer, E. W. & de Brabander-van den Berg, E. E. M. The dendritic box: shape-selective liberation of encapsulated guests. *J. Am. Chem. Soc.* **117**, 4417–4418 (1995).
- Maury, E. E. *et al.* Graft copolymer surfactants for supercritical carbon dioxide applications. *Polym. Prepr. (ACS Div. Polym. Chem.)* **34**, 664–665 (1993).
- McFann, G. J., Johnston, K. P. & Howdle, S. M. Solubilization in nonionic reverse micelles in carbon dioxide. *Am. Inst. Chem. Eng. J.* **40**, 543–555 (1994).
- Johnston, K. P. *et al.* Water-in carbon dioxide microemulsions: an environment for hydrophiles including proteins. *Science* **271**, 624–626 (1996).
- Reichardt, C. Empirical parameters of solvent polarity as linear free-energy relationships. *Angew. Chem. Int. Edn Engl.* **18**, 98–110 (1979).
- Zacharasee, K. A., Phuc, N. V. & Kozankiewicz, B. Investigation of micelles, microemulsions, and phospholipid bilayers with the pyridinium n-phenolbetaine $\text{E}_t(30)$, a polarity probe for aqueous interfaces. *J. Phys. Chem.* **85**, 2676–2683 (1981).
- Wignall, G. D. *et al.* Reduction of parasitic scattering in SAXS by a three-pinhole collimating system. *J. Appl. Crystallogr.* **23**, 241–246 (1990).
- Pfund, D. M., Zemanian, T. S., Linehan, J. C., Fulton, J. L. & Yonker, C. R. Fluid structure in supercritical xenon by nuclear magnetic resonance spectroscopy and small angle X-ray scattering. *J. Phys. Chem.* **98**, 11846–11857 (1994).
- Russell, T. P., Lin, J. S., Spooner, S. & Wignall, G. D. Intercalibration of small angle X-ray and neutron scattering data. *J. Appl. Crystallogr.* **21**, 629–638 (1988).

Acknowledgements. A.I.C. thanks the Royal Commission for The Exhibition of 1851 for a research fellowship (1995–96); J.D.L. thanks the NSF for support through the University of Tennessee. We also acknowledge support from the NSF for a Presidential Faculty Fellowship (J.M.D., 1993–97) as well as from the Consortium on Synthesis and Processing of Polymeric Materials in CO_2 , which is sponsored by the NSF, the Environmental Protection Agency, DuPont, Hoechst-Celanese, Air Products and Chemicals, B. F. Goodrich, Eastman Chemical, Bayer and Xerox. We also thank the Division of Material Sciences, US Department of Energy under contract with Lockheed-Martin Research Corporation, and the Exxon Education Foundation for support.

Correspondence and requests for materials should be addressed to J.M.D. (e-mail: desimone@unc.edu).

An Earth-like numerical dynamo model

Weijia Kuang & Jeremy Bloxham

Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA

The mechanism by which the Earth and other planets maintain their magnetic fields against ohmic decay is among the longest standing problems in planetary science. Although it is widely acknowledged that these fields are maintained by dynamo action, the mechanism by which the dynamo operates is in large part not understood. Numerical simulations of the dynamo process in the Earth's core^{1–4} have produced magnetic fields that resemble the Earth's field, but it is unclear whether these models accurately represent the extremely low values of viscosity believed to be appropriate to the core. Here we describe the results of a numerical investigation of the dynamo process that adopts an alternative approach⁵ to this problem in which, through the judicious choice of boundary conditions, the effects of viscosity are rendered unimportant. We thereby obtain a solution that at leading order operates in an Earth-like dynamical regime. The morphology and evolution of the magnetic field and the fluid flow at the core–mantle boundary are similar to those of the Earth, and the field within the core is qualitatively similar to that proposed on theoretical grounds⁶.

The dynamical regime of the geodynamo is most simply understood by considering the momentum balance in the Earth's fluid outer core. In addition to pressure gradients which play an essentially passive role, the leading order force balance is between the Coriolis force, the Lorentz force and the buoyancy force. Taylor⁷ showed that the magnetic field must then adjust so that the axial Lorentz torque on the surface of any cylinder in the core coaxial with the rotation axis (the so-called Taylor torque) is small, because the only torques available to balance it arise from the viscous force and from inertia, both of which are small. In the so-called magnetostrophic limit, the effects of inertia and viscosity vanish, and the Taylor torques must then be zero; in this limit the dynamo is described as being in a Taylor state.

In the Earth, small deviations from the Taylor state result in torsional oscillations in which the Taylor torque is balanced to leading order by inertia, with viscous effects most probably playing only a very small role. This regime is not easily modelled numerically, because the extremely small values of the viscosity that are required to reach such a quasi-Taylor-state result in viscous boundary layers far thinner than those that can be resolved numerically. In the Earth, the viscous boundary layer thickness lies in the range from 3 cm to 100 m, depending on whether the primary mechanism of viscous dissipation is molecular or turbulent, which is very small compared to the radius of the core of almost 3,500 km.

This view of the force balance in the geodynamo is supported by observations that suggest that changes in the angular momentum of the core are well described by torsional oscillations about a Taylor state^{8–10}. An alternative¹¹ is that the residual Taylor torques are balanced by viscous stresses at the boundaries. However, that model implies that angular momentum transfer between the outer core and mantle occurs on a timescale much longer than that suggested by observations.

Glatzmaier and Roberts^{1–4} have developed a numerical dynamo model which produces a predominantly dipolar field, similar to the Earth's field, and a westward drift of the field, at a rate similar to the Earth's field. However, it is unclear whether viscous effects are sufficiently small in their model, for not only is the value that they

choose for the viscosity much larger than that which is appropriate for the Earth, but viscous effects are further increased by the adoption of a hyperdiffusivity, whereby in order to achieve satisfactory numerical convergence, the diffusivities (viscous, magnetic, thermal and compositional) are increased with wavenumber, and by a further imposed increase in the viscosity by a factor of ten near the boundaries. The viscous boundary layers are then at least several orders of magnitude thicker than in the Earth's fluid outer core, and any inertial oscillations about a Taylor state would most likely be damped (by viscous dissipation) on a timescale that would be short compared to the oscillation period, with inertia thus not playing a significant part in balancing the Taylor torques.

Our approach, described in detail elsewhere⁵, is aimed specifically at obtaining a quasi-Taylor-state solution in which viscous effects are small. To reduce the role played by viscous effects, we impose viscous stress-free boundary conditions at the rigid boundaries. Although any viscous fluid must necessarily satisfy a no-slip boundary condition at a rigid boundary, we believe that it is a better approximation to the dynamical regime of the Earth's core, given the computational constraints, to impose a stress-free condition, thereby effectively reducing the boundary layer thickness to zero, because it is unlikely that the thin viscous boundary layers in the Earth's core are important in generating a planetary scale magnetic field, which is probably associated

Figure 1 A segment of a magnetic field line illustrating the dynamo process. The outer spherical surface with the latitude-longitude grid is the core-mantle boundary and the inner red surface is the inner-core boundary; the rotation axis is the straight line passing through the inner core. The field line is colour-coded so that it is yellow where it has a positive radial component and blue where it has a negative radial component. The field-generating mechanism can be visualized as follows. Starting at the top on the large loop of field protruding from the core, and tracing down and to the right, the field line enters the core, at the rear of the core in this view. Then, the field line is stretched around the rotation axis by the differential rotation (the mechanism by which the strong toroidal field is generated). Then, the effects of the poloidal part of the flow on the field can be seen, as the field line is twisted and spirals downwards almost parallel to the rotation axis. The effect of this helical distortion of the field line is apparent from the alternation between blue and orange along the field line as it is twisted. Next, the field line is again stretched by the differential rotation, before re-emerging from the core on the right-hand side. The field line then loops outside the core before re-entering the core (near the top right). On re-entering the core, this cycle repeats. This classical picture of generation of toroidal field by differential rotation and generation of poloidal field by helical motions was first postulated more than 40 years ago⁶. Starting again from the original starting point, but following the large loop in the other direction, we can see the field line penetrate the inner core.

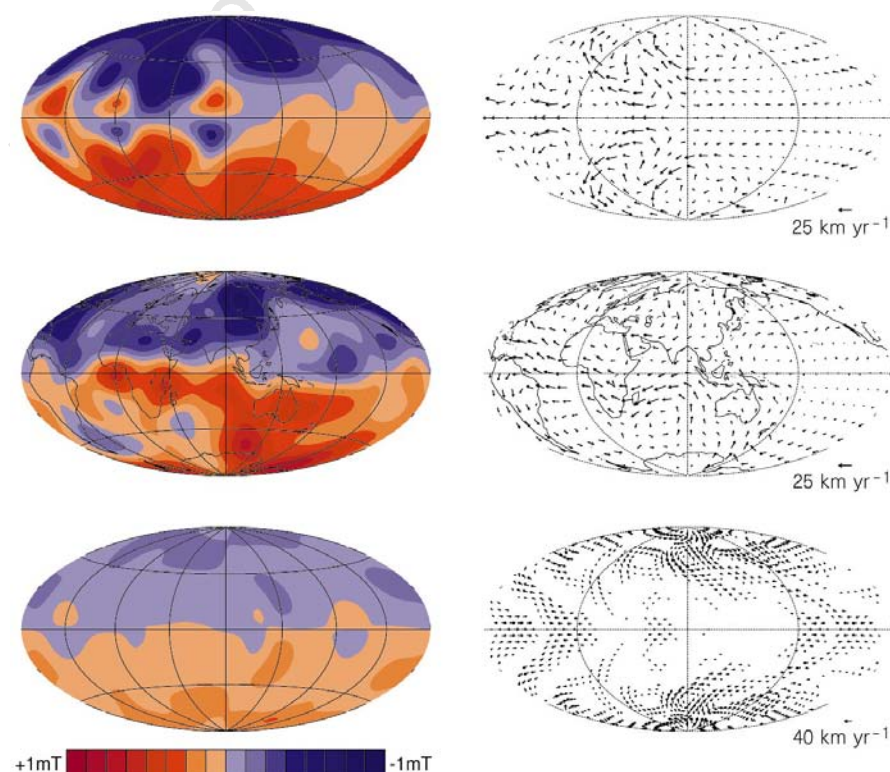
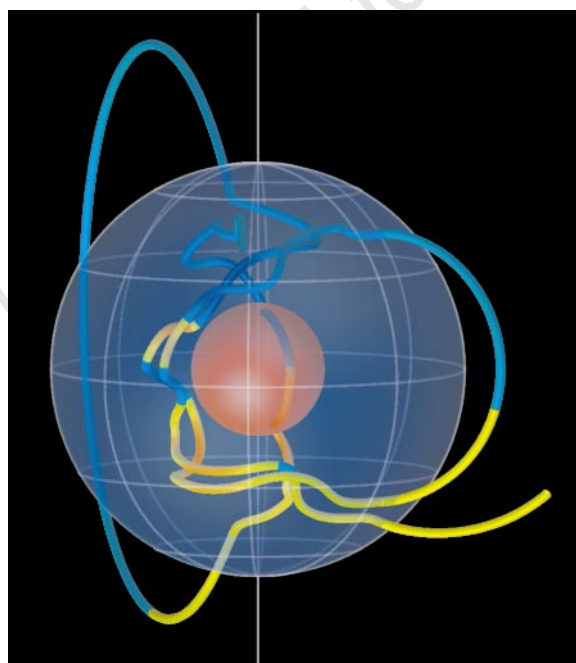


Figure 2 The radial component of the magnetic field (left column) and the fluid flow (right column) at the core-mantle boundary. The top row is from our dynamo model; the middle row shows the Earth's field and fluid flow averaged over the period 1840-1990^{14,15}; the bottom row is from the model of Glatzmaier and Roberts. Both dynamo models generate magnetic fields that have a basically dipolar structure like the Earth's field, although the Glatzmaier-Roberts model is only 30% of the strength of the geomagnetic field. Our model has a flow at the core-mantle boundary that is very similar to that of the Earth, with large gyres in each hemisphere extending close to the Equator forming a westward jet of about 180° extent in longitude. Elsewhere the flow is weak. The Glatzmaier-Roberts model (shown about 70 km beneath the core-mantle boundary) has a more complicated morphology, with strong eastward and westward jets near the Equator.

with the mainstream flow.

Assuming that turbulent diffusivities are appropriate for the Earth's core, we have chosen the non-dimensional numbers that define the system so that those that should be order one are of that order, and those that should be small are small, although, owing to numerical constraints, not as small as they should be. Of primary concern are the Ekman number and the magnetic Rossby number. We choose these each to have the value 2×10^{-5} ; by assigning them the same value, the magnetic Prandtl number (the ratio of viscous diffusivity to magnetic diffusivity) is unity. In the Earth's core, even with turbulent values for the diffusivities, these two numbers are several orders of magnitude smaller; and, like Glatzmaier and Roberts, we too enhance the diffusivities further at high wave-number, although using a weaker functional dependence on wave-number. The flow in our model is driven thermally at a Rayleigh number roughly 30 times the critical value for the onset of convective instability in the absence of a magnetic field.

Our model has maintained a strong magnetic field for roughly seven free decay times, or approximately 100,000 years. The solution was very slow to evolve from the initial starting state, taking about three decay times for the transients to decay fully. During this transient period, the magnetic field reversed polarity completely and underwent several near reversals; since then the polarity has been stable. The magnetic field is strong both in an energetic sense in that the magnetic energy is several orders of magnitude larger than the kinetic energy, and in a dynamical sense in that the resulting Lorentz force is important in the momentum balance at leading order. Our model is close to a Taylor state, with the magnetic field organized such that the Taylor torques are much smaller than would be estimated simply from the strength of the magnetic field in the core, which suggests that the effects of viscosity and inertia are not important at leading order. The process of magnetic field generation is shown in Fig. 1.

In Fig. 2, we show the magnetic field and the fluid flow at the core-mantle boundary for our model and, for comparison, for the Earth and for the Glatzmaier-Roberts model. In Fig. 3 we show the axisymmetric part of the magnetic field and the flow within the core from our model and from the Glatzmaier-Roberts dynamo model. From these two figures it is clear that, although each model produces a predominantly dipolar field at the core-mantle boundary (but of differing strength), the mechanism by which that field is generated and maintained is quite different. The distribution of toroidal field and the pattern of rotation and poloidal flow are markedly different.

Of the explanations that can be proposed for this difference between the models, the most reasonable is that it is due to the different force balance that results from the different boundary conditions imposed on the flow at the rigid boundaries. The no-slip boundary conditions with their concomitant large viscous stress results in a model that operates in a different dynamical regime from our model.

To test this, we changed the boundary conditions on the flow in our model to no-slip boundary conditions. We kept the same values of the non-dimensional parameters that we used in our original model in order to guard against the alternative possibility that the difference is due to different values of the non-dimensional numbers. After imposing these changes, our solution evolved over one decay time to a state much closer to the Glatzmaier-Roberts dynamo (Fig. 3).

This experiment indicates that the difference in these models is due primarily to the effect of viscous coupling on the flow, and that it is not due to different choices of the non-dimensional parameters. It also suggests that the generation of strong field near the inner-core boundary and the pattern of fluid flow in the Glatzmaier-Roberts dynamo is a consequence of the thick viscous boundary layer which occurs there in their model, but which does not occur in the Earth. In the Earth it is more likely that the role of the inner core is to

modulate the behaviour of the magnetic field^{12,13}. On the other hand, our model with viscous stress-free boundary conditions is unaffected by viscosity at leading order.

Finally, we have found that the differential rotation of the inner core is sensitive to the viscous boundary conditions. In our model,

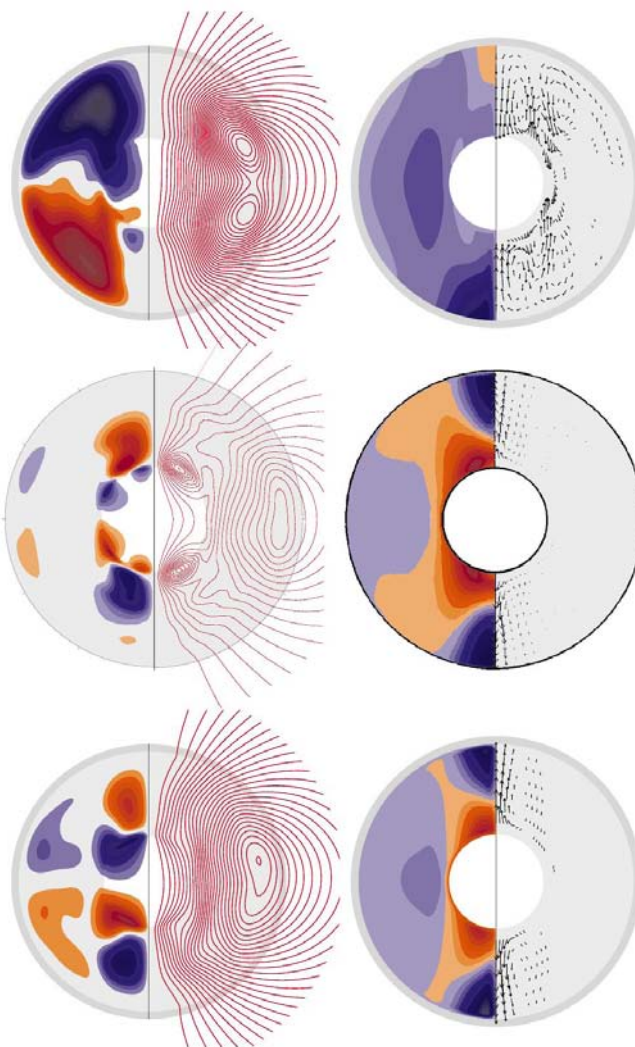


Figure 3 The axisymmetric parts of the magnetic field and the fluid flow in the core. Cross-sections through the core are shown, with the North Pole at the top and the South Pole at the bottom. The white innermost region is the solid inner core; the light grey region is the fluid outer core; and, where applicable, the darker grey layer represents an electrically conducting region at the base of the solid mantle. Top, our model; middle, the Glatzmaier-Roberts model; bottom, our model after applying no-slip boundary conditions. The left-hand column shows the magnetic field and the right-hand column the fluid flow. For plots of the magnetic field, the left-hand part shows the toroidal part of the field, with orange shades representing eastward-directed field and blue shades westward-directed field. The right-hand part shows field lines of the poloidal field. For the plots of fluid flow, the left-hand part shows the rotation of the fluid, with blue representing westward rotation relative to the mantle and orange representing eastward rotation. On the right, the arrows depict the poloidal part of the flow. In our model, the toroidal field has a simple structure, with strong field throughout the outer core, whereas in the Glatzmaier-Roberts model it is concentrated near the inner-core boundary. The poloidal fields also differ: in our model, it has a very simple structure whereas in the Glatzmaier-Roberts model it has the extra complication of strong closed loops of field near the inner core boundary. Likewise, the flow in our model is distributed throughout the core, instead of being concentrated near the rotation axis in the Glatzmaier-Roberts model. The bottom row shows that after applying no-slip boundary conditions, our model evolved to a state similar to the Glatzmaier-Roberts dynamo.

we observe periods of both eastward and westward differential rotation of the inner core, with a period of oscillation of several thousand years. After we imposed the viscous no-slip boundary condition, and following an initial transient adjustment, we have observed only eastward differential rotation of the inner core, suggesting that viscous effects may also be important in the dynamics of the inner core in the Glatzmaier–Roberts dynamo model⁴, probably indirectly through their strong role in determining the flow in the outer core, as suggested by the flow patterns in Fig. 3.

Much remains to be done to understand the geodynamo. In particular, we need to undertake a systematic exploration of the effect of varying the non-dimensional parameters in order to characterize the dynamics of dynamo action. Increasing our understanding of the dynamo process might allow us to develop a simplified model in which, for example, non-axisymmetric effects could be parametrized, thereby permitting numerical studies of the field evolution on much longer timescales. □

Received 12 March; accepted 15 July 1997.

- Glatzmaier, G. A. & Roberts, P. H. A three-dimensional convective dynamo solution with rotating and finitely conducting inner core and mantle. *Phys. Earth Planet. Inter.* **91**, 63–75 (1995).
- Glatzmaier, G. A. & Roberts, P. H. A three-dimensional self-consistent computer simulation of a geomagnetic field reversal. *Nature* **377**, 203–209 (1995).
- Glatzmaier, G. A. & Roberts, P. H. An anelastic evolutionary geodynamo simulation driven by compositional and thermal convection. *Physica D* **97**, 81–94 (1996).
- Glatzmaier, G. A. & Roberts, P. H. Rotation and magnetism of Earth's inner core. *Science* **274**, 1887–1890 (1996).
- Kuang, W. & Bloxham, J. Numerical modelling of magnetohydrodynamic convection in a rapidly rotating spherical shell I: Weak and strong field dynamo action. *J. Comp. Phys.* (submitted).
- Bullard, E. C. & Gellman, H. Homogeneous dynamos and terrestrial magnetism. *Phil. Trans. Soc. Lond. A* **247**, 213–278 (1954).
- Taylor, J. B. The magnetohydrodynamics of a rotating fluid and the Earth's dynamo problem. *Proc. R. Soc. Lond. A* **274**, 274–283 (1963).
- Jault, D., Gire, C. & Le Mouél, J.-L. Westward drift, core motions and exchanges of angular momentum between core and mantle. *Nature* **333**, 353–356 (1988).
- Jackson, A., Bloxham, J. & Gubbins, D. in *Dynamics of the Earth's Deep Interior and Earth Rotation* Vol. 72 (eds Le Mouél, J.-L., Smylie, D. & Herring, T.) (Geophys. Monog., Am. Geophys. Union, 1993).
- Zatman, S. & Bloxham, J. Torsional oscillations and the magnetic field within the Earth's core. *Nature* **388**, 760–763 (1997).
- Braginsky, S. I. An almost axially symmetrical model of the hydromagnetic dynamo of the earth, I. *Geomag. Aeron.* **15**, 149–156 (1975).
- Hollerbach, R. & Jones, C. A. A geodynamo model incorporating a finitely conducting inner core. *Phys. Earth Planet. Inter.* **75**, 317–327 (1993).
- Hollerbach, R. & Jones, C. A. Influence of the Earth's inner core on geomagnetic fluctuations and reversals. *Nature* **365**, 541–543 (1993).
- Bloxham, J. & Jackson, A. Time-dependent mapping of the magnetic field at the core–mantle boundary. *J. Geophys. Res.* **97**, 19537–19563 (1992).
- Bloxham, J. The steady part of the secular variation of the Earth's magnetic field. *J. Geophys. Res.* **97**, 19565–19579 (1992).

Acknowledgements. We thank G. Glatzmaier and P. Roberts for many helpful discussions regarding their work and for providing the data sets that were used to prepare Figs 2 and 3, and P. Olson for critically reviewing the manuscript. This work was supported by the David and Lucile Packard Foundation and by the NSF.

Correspondence should be addressed to J.B.

Acute stimulation of glucose metabolism in mice by leptin treatment

Seika Kamohara*§, Rémy Burcelin†§, Jeffrey L. Halaas*, Jeffrey M. Friedman*‡ & Maureen J. Charron†

*Laboratory of Molecular Genetics, and ‡Howard Hughes Medical Institute, the Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

†Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA

§These authors contributed equally to this work.

Leptin is an adipocyte hormone that functions as an afferent signal in a negative feedback loop regulating body weight^{1–4}, and acts by interacting with a receptor in the hypothalamus and other tissues^{5,6}. Leptin treatment has potent effects on lipid metabolism, and leads to a large, specific reduction of adipose tissue mass after

several days^{1,4}. Here we show that leptin also acts acutely to increase glucose metabolism, although studies of leptin's effect on glucose metabolism have typically been confounded by the weight-reducing actions of leptin treatment, which by itself could affect glucose homeostasis^{1–3}. We have demonstrated acute *in vivo* effects of intravenous and intracerebroventricular administrations of leptin on glucose metabolism. A five-hour intravenous infusion of leptin into wild-type mice increased glucose turnover and glucose uptake, but decreased hepatic glycogen content. The plasma levels of insulin and glucose did not change. Similar effects were observed after both intravenous and intracerebroventricular infusion of leptin, suggesting that effects of leptin on glucose metabolism are mediated by the central nervous system (CNS). These data indicate that leptin induces a complex metabolic response with effects on glucose as well as lipid metabolism. This response is unique to leptin, which suggests that new efferent signals emanate from the CNS after leptin treatment.

The effects of leptin on glucose metabolism were studied after intravenous (IV) or intracerebroventricular (ICV) infusions into wild-type mice. Three groups were analysed: a control group, which was treated with phosphate buffered saline (PBS) both IV and ICV; an IV-leptin group was infused with leptin IV and PBS ICV; and an ICV-leptin group received leptin ICV and PBS IV (Fig. 1). The cannulation of the C57BL/6J wild-type mouse third ventricle and the insertion of Alzet osmotic pumps were performed one week before implantation of an intravenous catheter. ICV infusion of a low dose of leptin (5 ng h⁻¹) was accomplished by changing the pump to a leptin-filled pump two days before experimental measurements, which is the time needed to fill the dead space in the tubing (see methods). When the ICV pumps were changed, indwelling catheters were implanted into the left femoral vein of the mice. Two days later, leptin (or PBS) was administered IV to freely moving mice as a 5-h intravenous infusion (1 µg h⁻¹). The IV leptin infusion began 2 h after food was removed (Fig. 1).

Several measures indicative of the state of glucose metabolism were made 3–5 h after leptin infusion was begun. When the infusion was completed (after a 7-h fast), the weight of treated and untreated mice was unchanged (Table 1). The plasma leptin concentration in the control mice was 3.0 ng ml⁻¹ (control), compared with 33.4 ng ml⁻¹ for the IV-leptin mice and 4.8 ng ml⁻¹ for the ICV-leptin mice. The plasma glucose, glucagon and insulin concentrations of the leptin-treated mice were not significantly different from those of the control mice (Table 1). However, although insulin levels were slightly lower in the leptin groups (0.8 ng ml⁻¹ for the IV-leptin mice and 1.1 ng ml⁻¹ for the ICV-leptin mice, against 1.5 ng ml⁻¹ for the controls), glucose turnover was significantly increased after leptin treatment (23.7 (IV-leptin), 21.8 (ICV-leptin) and 15.0 mg per kg per min⁻¹ (control) (Fig. 2, Table 1)).

These data suggest that leptin regulates both glucose output and glucose uptake, independent of increases in plasma insulin. A significant decrease in glycogen content in the liver (93.2 ± 7.0 (control), 59.5 ± 5.2 (IV-leptin), 63.2 ± 10.6 per (ICV-leptin); Table 1) was observed after leptin treatment, suggesting that glucose output from the liver was increased in the leptin-treated animals. The surgical manipulation resulted in a slight decrease in liver glycogen relative to fasted animals (109.8 ± 6.7 versus 93.2 ± 7.0 µg per mg), but this difference was not significant. An increased content of glycogen in skeletal muscle was also observed in the ICV-leptin group (1.4 ± 0.1 versus 2.4 ± 0.1 µg per mg; Table 1).

Glucose uptake into several tissues was measured by administering ¹⁴C-2-deoxyglucose. A significant increase in glucose uptake was observed in skeletal muscle and brown adipose tissue in the leptin-treated mice (Fig. 3a). Small increases in glucose uptake were observed in other tissues, but these were not significant. Consistent with this, leptin also resulted in a twofold increase in the whole-animal rate of glycolysis as scored by the conversion of D-3-³H glucose to ³H water (10.2 ± 1.6 (control) versus 17.3 ± 2.3 (IV-leptin)

Figure 1 Experimental protocols using C57BL/6J +/- lean mice. An ICV catheter was implanted at day -7. At day -2, the ICV osmotic pump was replaced with either leptin or PBS. At the same time, the femoral vein was catheterized and the leg denervated. Leptin was infused either IV or ICV for 5 h 40 min. After 3 h equilibration of leptin infusion, a primed continuous infusion of D-3-³H-glucose was performed to assess glucose turnover. Blood was sampled and analysed as described in the Methods. To assess individual tissue glucose uptake, a flash injection of U-¹⁴C 2-deoxyglucose was performed.

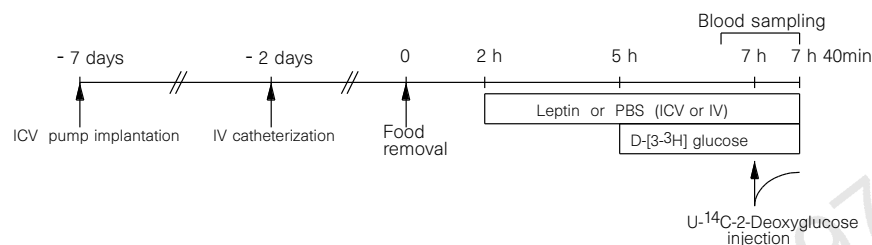


Table 1 Characteristics of leptin-treated and control mice

	Control	IV-leptin	ICV-leptin
No. of mice	8	7	6
Body weight (g)	25.4 ± 0.5	25.7 ± 0.8	26.6 ± 0.9
Basal plasma glucose (mg dl ⁻¹)	135 ± 11	124 ± 11	125 ± 13
Plasma glucose after 5-h fast (mg dl ⁻¹)	68 ± 6	61 ± 5	55 ± 4
Plasma glucose after treatment (mg dl ⁻¹)	70 ± 8	61 ± 5	51 ± 3
Lactate (mg dl ⁻¹)	70.1 ± 5.0	63.8 ± 8.0	82.6 ± 3.3
Plasma free fatty acids (μEq l ⁻¹)	547.1 ± 61.5	599.9 ± 53.0	511.1 ± 15.3
β-Hydroxybutyrate (mg dl ⁻¹)	4.7 ± 0.7	3.6 ± 0.7	3.6 ± 0.3
Leptin (ng ml ⁻¹)	3.0 ± 0.5	33.4 ± 6.0	4.8 ± 0.5
Insulin (ng ml ⁻¹)	1.5 ± 0.4	0.8 ± 0.2	1.1 ± 0.2
Glucagon (pg ml ⁻¹)	254.4 ± 44.2	282.1 ± 52.9	252.2 ± 34.4
T3 (ng dl ⁻¹)	52.3 ± 9.4	62.7 ± 12.1	102.5 ± 20.8*
T4 (μg dl ⁻¹)	0.7 ± 0.2	0.5 ± 0.1	0.3 ± 0.1*
Glucose turnover (mg per kg per min)	15.0 ± 1.2	23.7 ± 1.5***	21.8 ± 1.0**
Liver glycogen (μg per mg)	93.2 ± 7.0	59.5 ± 5.2*	63.2 ± 10.6*
Glycolysis (mg per kg per min)	10.2 ± 1.6	17.3 ± 2.3*	20.5 ± 0.2**
Glucose uptake (ng per mg per min)			
EDL	1.57 ± 0.43	4.13 ± 1.03*	6.46 ± 1.79*
Denervated EDL	3.05 ± 0.77	3.87 ± 0.75	3.87 ± 0.87
Soleus	1.86 ± 0.56	4.32 ± 1.22*	11.43 ± 2.48*
Denervated soleus	3.86 ± 1.07	8.35 ± 2.39	6.02 ± 1.39
Hindlimb glycogen (μg per mg)	1.4 ± 0.1	1.9 ± 0.3	2.4 ± 0.1*

We analysed 6–8 C57BL/6J wild-type mice per group. Plasma glucose was assessed before, during and after leptin and/or PBS infusion. Plasma hormone and metabolite levels were assessed at the completion of the infusions. The rate of glycolysis and glucose turnover were calculated as described in the Methods. EDL and soleus U-¹⁴C 2-deoxyglucose uptake was analysed from intact and denervated legs, and glycogen content was analysed from non-denervated leg on completion of study. Asterisks indicate statistical significance: (**P* < 0.05, ***P* < 0.005, ****P* < 0.001) between IV or ICV leptin-infused mice and PBS-infused control mice.

or 20.5 ± 0.2 mg per kg per min (ICV-leptin); Table 1), indicating that most of the labelled glucose was oxidized rather than being used as a substrate for lactate, glycogen or lipid biosynthesis.

We found that low-dose infusion of leptin ICV resulted in the same range of effects as seen with IV administration, but with greater potency. This finding suggests that the efferent signals from the CNS that modulate glucose metabolism are activated by leptin. The possibility that neuronal signals are responsible for the effect of ICV leptin on glucose uptake was tested by comparing glucose uptake in denervated and intact muscle in the same animal. Leptin-stimulated glucose uptake was decreased in denervated leg in both extensor digitorum longus (EDL, fast-twitch glycolytic fibres) and soleus (slow-twitch oxidative fibres) relative to the intact muscle (Fig. 3b, Table 1). These data suggest that the effects of leptin on glucose metabolism in skeletal muscle are dependent, in part, upon neuronal signals.

Leptin was unable to stimulate glucose uptake by EDL when the leg was denervated. However, glucose uptake in denervated soleus muscle after leptin treatment, although lower than that of the intact leg, was greater than in PBS-treated controls (Fig. 3b, Table 1). This residual increase in glucose uptake by denervated soleus could be the result of hormonal factors. Glucose uptake can also be modulated by thyroid hormone, which was measured in leptin-treated animals. Triiodothyronine (T3) levels were slightly higher (62.7 versus 52.3 ng dl⁻¹) and thyroxine (T4) levels were lower (0.5 versus 0.7 μg dl⁻¹) in the IV-leptin group than in PBS-treated controls, although those data were not significant. ICV leptin administration significantly increased the level of T3 (102.5 versus 52.3 ng dl⁻¹, *P* < 0.05) and decreased T4 (0.3 versus 0.7 μg dl⁻¹, *P*, 0.05), suggest-

ing that there was increased de-iodination of T4 to T3 rather than a net increase in production.

Several lines of evidence have suggested that leptin may modulate glucose metabolism^{1–3}. Leptin treatment of genetically obese *ob/ob* mice improves diabetes at doses below those affecting weight loss^{1–3}. Leptin treatment of lean mice does not change blood glucose, but a decrease in hepatic glycogen has been observed⁷. Recent studies using cultured hepatoma cells have also suggested that leptin may inhibit insulin action *in vitro*⁸. Finally, insulin has been shown to alter leptin gene expression⁹. In their studies, however, the effect of

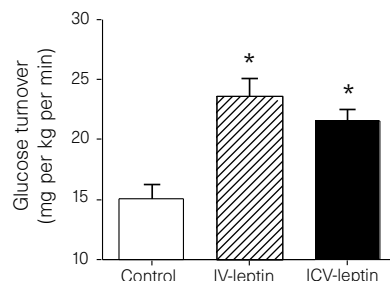


Figure 2 Glucose turnover in C57BL/6J +/- lean mice after IV or ICV leptin infusion. Glucose turnover was assessed after infusions of leptin or PBS. Leptin increased glucose turnover from 15.0 ± 1.2 to 23.7 ± 1.5 mg per kg per min (*P* < 0.001) with IV, or to 21.8 ± 1.0 mg per kg per min (*P* < 0.005) with ICV treatment. Asterisks indicate statistical significance between leptin-treated and control mice; 6–8 mice per group were studied.

leptin on glucose metabolism *in vivo* have not been carefully evaluated.

Our data show that leptin affects glucose metabolism. Although plasma glucose levels did not change in response to leptin, glucose uptake and oxidation were increased. These data suggest that, in addition to its effects on lipid metabolism, the negative energy balance in leptin-treated animals is also associated with increased metabolism of glucose. The increased glucose turnover in the absence of a significant change in insulin concentration suggests that the leptin-mediated increase in glucose metabolism is not likely to be the result of an increased rate of insulin secretion in C57BL/6J $+/+$ mice. It could be the result of an increase in insulin sensitivity, although it remains possible that leptin treatment results in increased glucose uptake through an insulin-independent mechanism. The latter possibility is suggested by the observed increase in the rate of glycolysis, and further studies are required to resolve this. The increase in glycolysis is also independent of glucagon, as the levels of this hormone were also unchanged after leptin treatment. The increase in glucose turnover while the plasma glucose level was unchanged suggests that leptin might also increase glucose output, perhaps through the liver. This idea is consistent with our data and with previous reports showing that chronic treatment with leptin decreases hepatic glycogen content⁷.

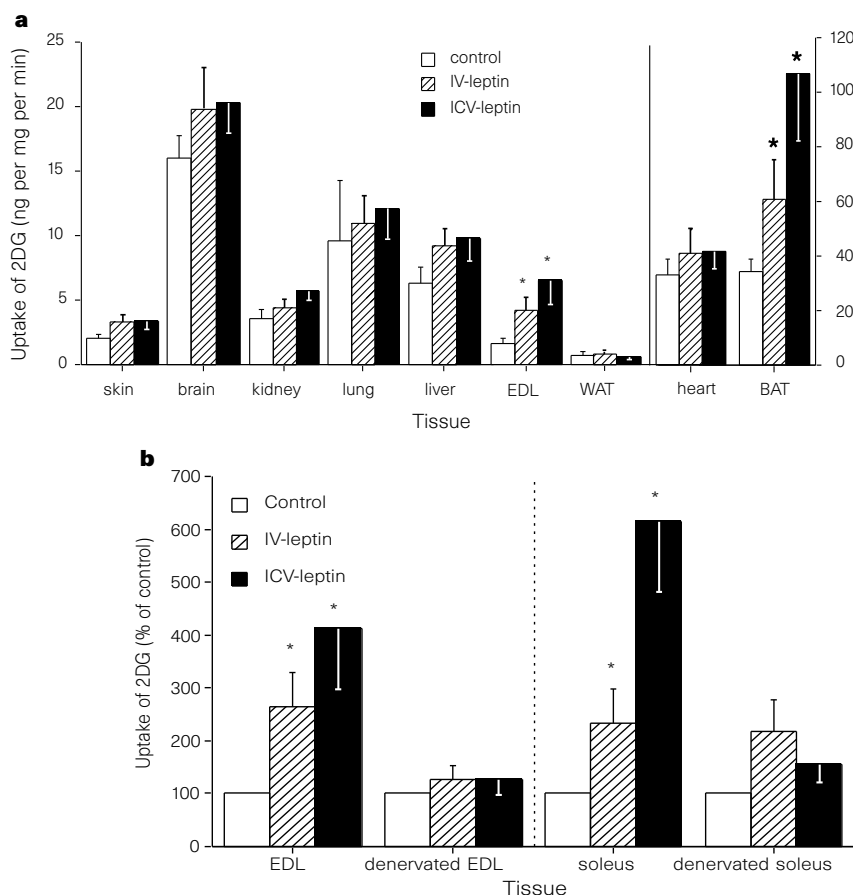
The data from mice receiving the IV leptin infusion indicate that leptin acutely alters glucose metabolism. Low-dose ICV infusion of leptin resulted in the same response as IV administration, suggesting that the observed effects of peripheral leptin on glucose metabolism are likely to be the result of transport into the CNS and interaction with receptors there^{5,6,10,11}. Several other lines of evidence suggest that the hypothalamus is an important target of both ICV and IV leptin. Leptin levels increase after hypothalamic lesions, and rats with lesions in the ventromedial hypothalamus do not respond to

exogenous leptin^{12,13}. The signalling form of the leptin receptor is highly expressed in those hypothalamic nuclei previously shown to control the regulation of food intake and body weight^{5,6,10,14}. Moreover, the arcuate, ventromedial hypothalamus and lateral hypothalamus (LH) signalling form of the receptor is not expressed to a great extent in any other brain region¹⁴. The activity of the Stat3 transcription factor and *fos* gene expression are increased within 1 h of a single dose of IV leptin¹⁵. Finally, the 100-fold higher potency in inducing weight loss of ICV leptin than subcutaneous leptin⁴ and the identification of leptin in cerebrospinal fluid are consistent with the main site of action of leptin being in the CNS^{16,17}.

The effects of ICV leptin on glucose uptake indicates that this response is the result of a stimulation of efferent pathways from the CNS. Numerous studies have suggested that the CNS, in particular the hypothalamus, can regulate glucose metabolism^{18–21}. These studies also suggest that alterations in sympathetic tone may mediate this effect²². The stimulatory effect of leptin on glucose output and glucose uptake that we observed is similar to that seen during exercise²³ or in response to ICV carbachol¹⁸. The possibility of an overlap between the stimulatory pathways mediating the action of leptin on carbohydrate metabolism and that of carbachol or exercise is intriguing, and requires further investigation.

Our data further suggest that some of the effects of leptin are mediated through neural signals. Previous data have shown that leptin treatment increases sympathetic output to brown fat through the activation of adrenergic receptors²². The data do not exclude the possibility that hormonal factors also have an effect. We have shown that ICV leptin leads to elevated T3 and decreased T4, suggesting that leptin treatment increases the rate of conversion of T4 to T3. The activity of the de-iodinase that converts T4 to T3 is also known to be modulated by the autonomic nervous system. In addition, thyroid hormone is known to increase basal glucose uptake by

Figure 3 Individual tissue glucose uptake in C57BL/6J $+/+$ lean mice after IV or ICV leptin infusion. **a**, U-¹⁴C 2-deoxyglucose (2-DG) uptake was assessed after IV or ICV infusions of leptin or PBS. The tissues tested included skin, brain, kidney, lung, liver, extensor digitorum longus (EDL), periovarian white adipose tissue (WAT), heart and inter-scapular brown adipose tissue (BAT). A significant increase (asterisks) in glucose uptake was observed in EDL ($P < 0.05$), and BAT ($P < 0.05$) in leptin-infused animals; 6–8 mice per group were studied. **b**, 2-DG uptake was assessed in intact and denervated EDL and soleus after leptin or PBS infusions. Asterisks indicate statistical significance with $P < 0.05$ between leptin- and PBS-treated mice; 6–8 mice per group were studied.



various tissues^{24,25}. Thus an increase in T3 levels could also account for a fraction of the leptin-induced increase in glucose uptake by other tissues. It has previously been shown that leptin can also increase T4 levels in fasted mice²⁶. The basis for this difference is unclear, although the length of the period of food restriction was different in these studies. The data do not exclude the possibility that other hormonal signals are also activated by leptin.

In summary, our *in vivo* studies indicate that acute administration of leptin increases glucose metabolism independent of elevations in plasma insulin by interacting with receptors in the CNS. These findings suggest that leptin induces a state of negative energy balance and increases metabolism of both fat and glucose. It is also possible that mechanisms to provide glucose to peripheral tissues have evolved during leptin-mediated anorexia. □

Methods

Experimental procedures. C57 BL6J *+/+* lean mice were anaesthetized with ketamine/xylazine. A 1-cm midline incision was made across the top of the skull, and the animal was placed on a stereotaxic apparatus and the periosteum cleaned. A hole 1 mm in diameter was made midline 0.3 mm behind the bregma, and a 30-gauge stainless steel infusion cannula was implanted into the third ventricle using the following coordinates: 0.3 mm relative to bregma, midline 3 mm ventral. Two supporting screws 0.5 mm in diameter were placed bilaterally in the posterior quadrants of the skull and the cannula secured in place with dental acrylic. The cannula was connected to 6 cm tygon tubing and an osmotic pump (Alzet, Palo Alto, CA) containing CSF placed at the other end of the tubing. The pump and tubing were placed in the subcutaneous space dorsally and the skin sutured. The mice were observed for 8 days, with weight, food intake and behaviour being monitored daily to verify a total recovery. The time required for peptide solutions to reach the third ventricle was estimated by infusing leptin or angiotensin, a peptide that results in a dramatic increase in water intake (J.L.H. *et al.*, unpublished data). The mice were anaesthetized 48 h before glucose metabolism was analysed and the osmotic pumps were replaced by new ones filled with PBS or leptin. At the same time an indwelling intravenous catheter was implanted. Briefly, a 5-mm vertical skin incision was made on a ventral side of the left leg at the level of the hip. A catheter was implanted into the vena cava through the femoral vein. A piece of the adjacent nerve was physically removed from the leg to insure denervation. The catheter was sealed subcutaneously, made exterior at the base of the neck, and glued in place with dental cement. There were three groups for this study: a control group was treated with PBS both IV and ICV; an IV-leptin group was infused with leptin IV and PBS ICV; and an ICV-leptin group received leptin ICV and PBS IV. To avoid glucose released from the digestive tract, after 48 h the animals were food-deprived for 2 h before the leptin or PBS infusions started (Fig. 1). Under these conditions the rate of glucose appearance into the blood represents the rate of hepatic glucose production. To assess glucose turnover, a prime (2.33 μ Ci) continuous infusion (7 μ Ci per kg per min) of HPLC-purified D-3-³H-glucose (Amersham) was maintained for 120 min in unrestrained mice. At periods of 80, 90, 100, 110 and 120 min of D-3-³H glucose infusion, 10 μ l of blood was sampled through the tail vein, deproteinized by the Somogyi procedure, and the specific activity of D-3-³H-glucose in the supernatant was analysed²⁷. To analyse individual tissue glucose uptake, 0.5 μ Ci g⁻¹ or ¹⁴C-2-deoxyglucose was flash-injected into the vein after the D-3-³H-glucose infusion. Tail blood was sampled at 2, 4, 6, 10, 20, 30 and 40 min after the injection. The animals were killed by cervical dislocation, the organs removed, and individual tissue glucose uptake calculated²⁸.

Plasma hormone and metabolite determinations. After the infusions, blood was drawn from the retro-orbital sinus using a heparinized micro-capillary tube and centrifuged. The plasma was collected immediately and stored at -70°C. Plasma insulin, thyroid hormones and leptin levels were measured using rat insulin (Linco, St Louis, MO), thyroxine hormones (DPC, Los Angeles, CA) or mouse leptin (Linco, St Louis, MO) radioimmunoassay kits, respectively. Plasma glucose, lactate and β -hydroxybutyrate were measured using the appropriate kits (Sigma, St Louis, MO). Plasma free fatty

acids were assessed (Amano, Richmond, VA) with oleic acid as standard. Muscle and liver glycogen content was determined^{28,29}.

Calculations. $GT = \text{gluIR}/\text{gluSA} = \text{HGP}$, where GT is glucose turnover (mg per kg per min), gluIR is glucose infusion rate (d.p.m.), gluSA is the specific activity of D-3-³H-glucose in the blood (d.p.m. mg⁻¹), and HGP is hepatic glucose production (mg per kg per min). Individual tissue glucose uptake and glycolysis rates were calculated^{28,29}. Results are presented as mean $\pm$ s.e. Statistical significance was determined by the Kruskal-Wallis test for muscle glucose uptake determinations and Mann-Whitney tests for significance between groups. The Student's *t*-test was used for other unpaired data.

Received 2 June; accepted 18 July 1997.

- Halaas, J. L. *et al.* Weight-reducing effects of the plasma protein encoded by the *obese* gene. *Science* **269**, 543–546 (1995).
- Campfield, L. A., Smith, F. J., Guisez, Y., Devos, R. & Burn, P. Recombinant mouse OB protein: evidence for a peripheral signal linking adiposity and central neural networks. *Science* **269**, 546–549 (1995).
- Pelkeymounter, M. A. *et al.* Effects of the *obese* gene product on body weight regulation in *ob/ob* mice. *Science* **269**, 540–543 (1995).
- Halaas, J. L. *et al.* Physiological response to long-term peripheral and central leptin infusion in lean and obese mice. *Proc. Natl Acad. Sci. USA* **94**, 8878–8883 (1997).
- Tartaglia, L. A. *et al.* Identification and expression cloning of a leptin receptor, OB-R. *Cell* **83**, 1263–1271 (1995).
- Lee, G. H. *et al.* Abnormal splicing of the leptin receptor in *diabetic* mice. *Nature* **379**, 632–635 (1996).
- Levin, N., Nelson, C., Gurney, A., Vanelen, R. & De Sauvage, F. Decreased food intake does not completely account for adiposity reduction after ob protein infusion. *Proc. Natl Acad. Sci. USA* **93**, 1726–1730 (1996).
- Cohen, B., Novick, D. & Rubinstein, M. Modulation of insulin activities by leptin. *Science* **274**, 1185–1188 (1996).
- Saladin, T. *et al.* Transient increase in *obese* gene expression after food intake or insulin administration. *Science* **377**, 527–529 (1995).
- Mercer, J. G. *et al.* Localization of leptin receptor mRNA and the long form splice variant (Ob-R) in mouse hypothalamus and adjacent brain regions by *in situ* hybridization. *FEBS Lett.* **387**, 113–116 (1996).
- Golden, P. L., Maccagnan, T. J. & Pardridge, W. M. Human blood-brain barrier leptin receptor. *J. Clin. Invest.* **99**, 14–18 (1997).
- Maffei, M. *et al.* Increased expression in adipocytes of *ob* RNA in mice with lesions of the hypothalamus and with mutations at the *db* locus. *Proc. Natl Acad. Sci. USA* **92**, 6957–6960 (1995).
- Satoh, N. *et al.* Pathophysiological significance of the *obese* gene product, leptin, in ventromedial hypothalamus (VMH)-lesioned rats: evidence for loss of its satiety effect in VMH-lesioned rats. *Endocrinology* **138**, 947–954 (1997).
- Fei, H. *et al.* Anatomic localization of alternatively spliced leptin receptors (Ob-R) in mouse brain and other tissues. *Proc. Natl Acad. Sci. USA* **94**, 7001–7005 (1997).
- Vaisse, C. *et al.* Leptin activation of Stat3 in the hypothalamus of wild type and *ob/ob* mice but not *db/db* mice. *Nature Genet.* **14**, 95–97 (1996).
- Schwartz, M. W., Seeley, R. J., Campfield, L. A., Burn, P. & Baskin, D. G. Identification of targets of leptin action in rat hypothalamus. *J. Clin. Invest.* **98**, 1101–1106 (1996).
- Caro, J. F. *et al.* Decreased cerebrospinal-fluid/serum leptin ratio in obesity: a possible mechanism for leptin resistance. *Lancet* **348**, 159–161 (1996).
- Miles, P. D. G., Yamatani, K., Lickley, H. L. A. & Vranic, M. Mechanism of glucoregulatory responses to stress and their deficiency in diabetes. *Proc. Natl Acad. Sci. USA* **88**, 1296–1300 (1991).
- Leong, S. F. & Clark, J. B. Regional enzyme development in rat brain: Enzymes associated with glucose utilization. *Biochemistry* **218**, 131–138 (1984).
- Lautala, P. & Martin, J. M. Glucose metabolism in rat hypothalamus. *Acta Endocrinol.* **98**, 481–487 (1981).
- Nagai, K., Fujii, T., Inoue, S., Takamura, Y. & Nakagawa, H. Electrical stimulation of the suprachiasmatic nucleus of the hypothalamus causes hyperglycemia. *Hormone Metab. Res.* **20**, 37–39 (1988).
- Collins, S. *et al.* Role of Leptin in fat regulation. *Nature* **380**, 677 (1996).
- Vranic, M., Kawamori, R., Pek, S., Kovacevic, N. & Wrenshall, G. A. The essentiality of insulin and the role of glucagon in regulating glucose utilization and production during strenuous exercise in dogs. *J. Clin. Invest.* **57**, 245–255 (1976).
- Gosteli-Peter, M. A., Schmid, C. & Zapf, J. Triiodothyronine increases glucose transporter isotype 4 mRNA expression, glucose transport, and glycogen synthesis in adult rat cardiomyocytes in long-term culture. *Biochem. Biophys. Res. Commun.* **221**, 521–524 (1996).
- Weinstein, S. P., O'Boyle, E. & Haber, R. S. Thyroid hormone increases basal and insulin-stimulated glucose transport in skeletal muscle. The role of GLUT4 glucose transporter expression. *Diabetes* **43**, 1185–1189 (1994).
- Ahima, R. S. *et al.* Role of leptin in the neuroendocrine response to fasting. *Nature* **382**, 250–252 (1996).
- Somogyi, M. Determination of blood sugar. *J. Biol. Chem.* **160**, 69–73 (1945).
- Tsao, T. S., Burcelin, R., Katz, E. B., Huang, L. & Charron, M. J. Enhanced insulin action due to targeted GLUT4 overexpression exclusively in muscle. *Diabetes* **45**, 28–36 (1996).
- Massillon, D., Barzilai, N., Hawkins, M., Prus-Wertherimer, D. & Rossetti, L. Quantitation of hepatic glucose fluxes and pathways of hepatic glycogen synthesis in conscious mice. *Am. J. Physiol.* **269**, E1037–E1043 (1995).

Acknowledgements. We thank E. B. Katz, C. Vaisse, J. Li and T. S. Tsao for discussions; J. Blaire-West and D. A. Denton for help with the ICV surgery; S. Korres for helping to prepare the manuscript; and Amgen for recombinant leptin. This work was supported by grants from the NIH (J.M.F. and M.J.C.), Pew Charitable Trust (M.J.C.), Juvenile Diabetes Foundation International (R.B.), the Philippe Foundation (R.B.) and the Manpei Suzuki Diabetes Foundation (S.K.).

Correspondence and requests for materials should be addressed to M.J.C. (e-mail: charron@aecon.yu.edu).

Retinal direction selectivity after targeted laser ablation of starburst amacrine cells

Shigang He & Richard H. Masland

Howard Hughes Medical Institute, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

Directionally selective retinal ganglion cells respond strongly when a stimulus moves in their preferred direction, but respond little or not at all when it moves in the opposite direction^{1,2}. This selectivity represents a classic paradigm of computation by neural microcircuits, but its cellular mechanism remains obscure. The directionally selective ganglion cells receive many synapses from a type of amacrine cell termed 'starburst' because of its regularly spaced, evenly radiating dendrites^{3,4}. Starburst amacrine cells have a synaptic asymmetry that has been proposed as the source of the directional response in the ganglion cells^{5,6}. Here we report experiments that make this unlikely, and offer an alternative concept of the function of starburst cells. We labelled starburst cells in living retinas, then killed them by targeted laser ablation while recording from individual directionally selective ganglion cells. Ablating starburst cells revealed no asymmetric contribution to the ganglion cell response. Instead of being direction discriminators, the starburst cells appear to potentiate generically the responses of ganglion cells to moving stimuli. The origin of direction selectivity probably lies with another type of amacrine cell.

Much pharmacological and electrophysiological evidence implicates starburst cells in the function of directionally selective ganglion cells^{4,7,8}, and some models propose that the starburst cells themselves are the elements that discriminate the direction of stimulus movement^{5,6}. These models postulate that an oriented subset of the dendrites of starburst cells is selectively connected to dendrites of the directionally selective ganglion cell. The dendrites of starburst cells both receive synaptic inputs and make synaptic outputs. Because of a known asymmetry between the inputs and outputs along the starburst cell's dendrites³, they could selectively stimulate the ganglion cell when a stimulus moves in the ganglion cell's preferred direction, or selectively inhibit it in the opposite (null) direction (Fig. 1). Either is possible because the starburst cells release both acetylcholine and GABA (γ -aminobutyric acid)^{9,10}.

To test this model, we photoablated starburst cells while recording from directionally selective ganglion cells. The recording was done in intact, isolated retinas that had previously been exposed to DAPI, which selectively accumulates in the nuclei of the starburst cells^{11,12}. Labelling with DAPI allowed the starburst cells to be identified in the living tissue (Fig. 2), and also helped restrict adventitious damage, because a laser could be tuned to DAPI's peak absorption wavelength. The laser method¹³ has been used for many studies in *Caenorhabditis elegans*¹⁴. A pulsed nitrogen laser was focused through the optics of the microscope to a spot 2 μ m in diameter centred within the cell body of a displaced starburst cell. From the optical geometry of the objective used, the density of laser energy at the level of the inner cells of the inner nuclear layer (the next DAPI-containing nuclei in the light path) is calculated to be approximately 0.5% of that at the level of the targeted starburst nucleus. This is an upper limit, because energy is subtracted from the beam during its passage through the DAPI-labelled nucleus of the targeted cell.

The somas of 20–60 displaced (on-type) starburst cells with dendrites that overlapped the receptive field of the ganglion cells were irradiated (Fig. 2c). When a starburst cell was irradiated, its soma became permeable to ethidium homodimer (Fig. 2b). The

disruption was evident within a few seconds and lasted for at least 3 h, as seen when a test pulse of ethidium was applied after that delay. Neighbouring cells beside or below the targeted starburst nucleus did not accumulate ethidium. After laser irradiation of dendrites in a neuron of the cricket¹⁵, the dendrites of a few (10%) of the cells resealed and the distal dendrite continued to function. However, that differs from our experiment, in which the cell soma was disrupted and did not resealed. When we irradiated the somas of cultured retinal neurons labelled by uptake of calcein, the fluorescent probe was depleted within seconds from their dendrites. Taken together, these findings indicate that laser irradiation specifically inactivated the entire targeted starburst cell. This was supported by the nature of the results (Fig. 3).

Ablating starburst cells on the null side of the receptive field (the side where stimulus movement in the null direction starts) had no

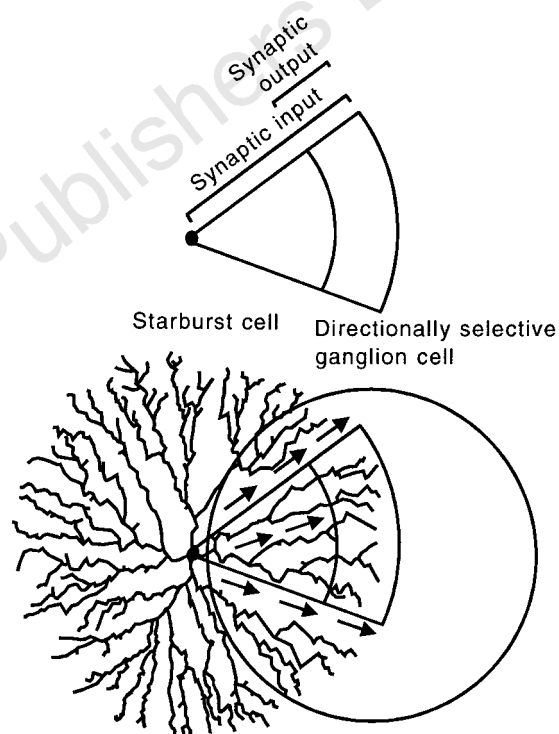


Figure 1 Starburst cells receive inputs from bipolar cells along the whole length of their dendrites, but their outputs to retinal ganglion cells are restricted to the distal third of the dendrites, leading to an asymmetric relationship between input and output. Some models propose that this is the anatomic asymmetry long recognized¹ as necessary for direction selectivity. A stimulus moving from proximal to distal along the starburst cell's dendrite will generate an output that tends to precede the arrival of the stimulus. Under certain assumptions about the properties of the dendrites⁶, their output is directionally selective. It is postulated that excitation within one sector of a starburst cell's dendritic arbor does not propagate to other sectors. Sectors that have the same retinal orientation could make synapses exclusively with a particular directionally selective ganglion cell. (For each directionally selective ganglion cell, the afferent starburst sectors would belong to many different starburst cells.) If the uniformly oriented sectors excite the ganglion cell, through acetylcholine, their orientation would define the ganglion cell's preferred direction. Alternatively, if they inhibit the ganglion cell, through GABA, they would define the null direction. Starburst sectors having other orientations would synapse with ganglion cells that have other preferred directions (see refs 5, 6).

effect on the responses of the ganglion cells. Neither the index of directionality nor the absolute value of the response was affected. This was not because the laser failed to inactivate the starburst cells, or because too few cells were targeted, as irradiating similar numbers of cells on the preferred side had a clear and irreversible effect (see below). If direction selectivity were caused by inhibition from starburst cells located on the null side of the receptive field, then ablating starburst cells on the null side should have released the inhibition and allowed a response to stimuli moving in the null direction. This would have been apparent for the leading (on) edge of the moving bar, because it was the 'on' starburst cells that were ablated. Instead, there was no discernible effect upon either the integrated responses (Fig. 3c, d) or the individual histograms (Fig. 3a). One would have expected at least a change in response when the leading edge of the stimulus crossed the maximally

deafferented side of the receptive field (see Fig. 1), but no such effect was observed. This is inconsistent with inhibition from the null-side starburst cells being a principal cause of directional selectivity.

In contrast, ablating starburst cells on the preferred side caused a depression of the responses of the ganglion cells. As expected, the effect was greatest for the 'on' component of the response and for the side of the ganglion cell's receptive field overlain by the ablated starburst cells (Fig. 3b). This reduction in the ganglion cell's response is consistent with the results of two-spot experiments, which showed that a facilitatory zone exists on the preferred side of the receptive field, outside the main receptive field¹⁶. We conclude that the preferred-side starburst cells enhance the directionally selective ganglion cell's response to moving stimuli.

The two-spot experiments were interpreted as supporting a starburst-based model of direction selectivity (Fig. 1). It was postulated that facilitation is created only by movement from the preferred side and is the fundamental mechanism that creates the directional discrimination^{6,16}. However, a different interpretation is possible. The primary cause of direction selectivity in the ganglion cell could be null-direction inhibition from another cell, as was argued in several earlier models^{8,17,18}. Starburst cells would facilitate responses to motion in all directions, but in the null direction that facilitation would be overwhelmed by the null-direction inhibition. In our experiments, further depression of the null response resulting from laser ablation of the null-side starburst cells would not be observable, because the response to movement in the null direction is virtually zero even under normal conditions. To distinguish between these alternatives, we performed two additional experiments.

The first investigated how tightly, if at all, cholinergic facilitation is restricted to the preferred direction. Many reports show that cholinergic antagonists depress the directionally selective ganglion cell's response to movement in the preferred direction but do not eliminate direction selectivity^{7,8,19–21}. We repeated that experiment, but collected complete direction tuning curves, to look for evidence of cholinergic facilitation in directions other than the preferred direction (Fig. 4). If excitatory starburst input occurred selectively from the preferred direction, then the contribution of cholinergic input should be restricted to the preferred direction. Instead, D-tubocurarine depressed the response at every angle where the response was large enough to be measured reliably. These included directions far from the preferred direction. For example, the response at 0 deg (the preferred direction) was depressed by 47%, but for the two directions 120 deg away—closer to the null direction than to the preferred—a substantial (36%) depression was also observed (Fig. 4). The index of directionality was virtually unchanged, from 0.94 ± 0.09 under control conditions to 0.87 ± 0.15 in D-tubocurarine. Starburst input thus enhances responses of the directionally selective cells over many directions, a result inconsistent with the straightforward model shown in Fig. 1. However, this finding does not eliminate the possibility of a selective 'hole' in the starburst input, aligned with the null direction.

To examine that possibility, the next experiment combined the use of two receptor antagonists. If facilitatory cholinergic input from the null side exists, we reasoned, it is masked by null-direction inhibition^{8,22}, and removing that inhibition would make the null-side starburst input observable. Null-direction inhibition can be removed by picrotoxin, resulting in a directionally selective ganglion cell that now responds to movement in all directions^{8,22}. In the presence of picrotoxin (Fig. 5), the index of directionality was reduced from 0.83 ± 0.12 to 0.08 ± 0.14 . When D-tubocurarine was applied, we observed a decrease in the response to movement in every direction, including the original (pre-picrotoxin) null direction. Indeed, the decrease in response was quite symmetrical, indicating a loss of starburst input from all sides of the receptive field.

Note that this last experiment relates to the question of starburst-mediated inhibition as well as starburst-mediated excitation. The

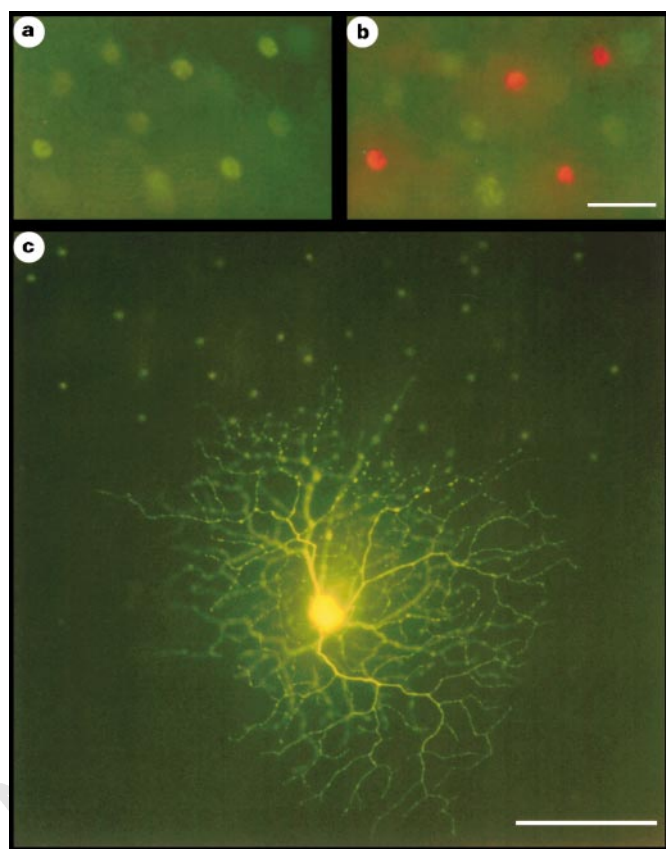


Figure 2 Photoablating on-starburst cells by laser irradiation. **a**, Living starburst cells in a retina labelled by exposure to DAPI. The starburst cells are uniform in size and are brighter than the surrounding ganglion cells. **b**, Four of the cells were irradiated for 5 s by a 365-nm pulsed laser focused to a spot 2 μ m in diameter. The irradiated cells became permeable to ethidium homodimer, which fluoresces at longer wavelengths and indicates cell death. The DAPI-labelled cells in **a** and **b** appear green rather than the usual blue because we used a broad band filter combination (peak excitation 425 nm, viewing >460 nm) that allows simultaneous visualization of DAPI and ethidium. Scale bar, 20 μ m. **c**, Ablation of starburst cells on one side of a directionally selective ganglion cell. The response properties of the ganglion cell were established using extracellular recording, then the somas of 30 starburst amacrine cells were photoablated. After irradiation, the cells become autofluorescent; they are visible here as bright spots scattered to one side of the dendritic arbor. The response properties of the ganglion cell were then studied again. After the physiological experiments had been completed, a Lucifer yellow-filled micropipette was used to fill the dendritic arbor of the directionally selective ganglion cell. The receptive fields of the directionally selective ganglion cells were ~10% larger than their dendritic arbors. The arbors of the nearest ablated starburst cells overlap the dendritic arbor of the ganglion cells by slightly more than 50% (ref. 12). Scale bar, 100 μ m.

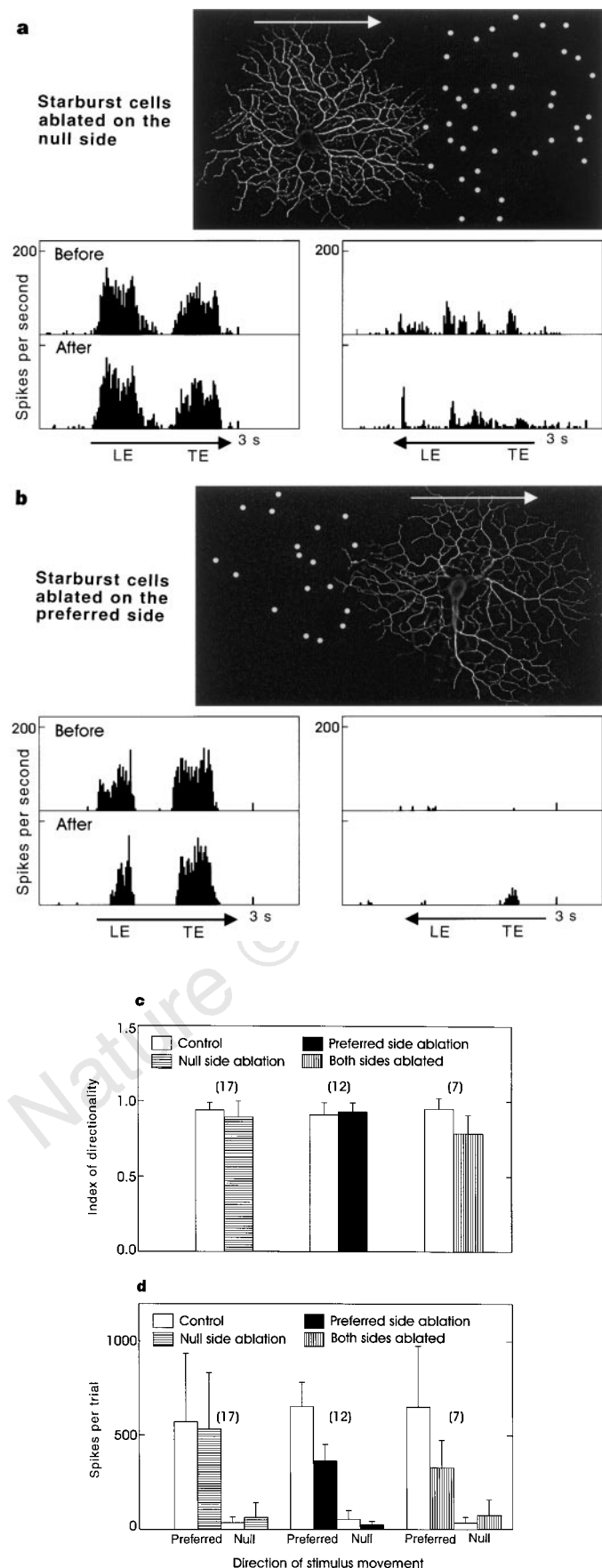


Figure 3 Effect of ablating starburst cells on the responses of directionally selective cells. The responses of the cell were studied before and after the ablations, and the ganglion cell was then injected with Lucifer yellow. The positions of the somas of the targeted starburst cells are shown by white dots. The dendritic arbors of the starburst cells are highly stereotyped; each is about 10% larger in diameter than the arbor of directionally selective cells at the same eccentricity^{12,28}. Poststimulus response histograms are means for 10 trials, and show two peaks, corresponding to the leading edge (LE) and the trailing edge (TE) of the moving rectangle. The direction of movement is indicated by arrows beneath the histograms. Ablating 39 cells on the null side of the cell shown in **a** had no discernible effect on the cell's response to a moving rectangle. Ablating 20 cells on the preferred side of the cell shown in **b** depressed the response to stimuli moving in the preferred direction. (Only 18 ablated cells are shown because the most distant two did not fit in the micrograph.) The effect was greatest for the initial part of the response, corresponding to that part of the receptive field overlain by the ablated starburst cells, and for the on (leading edge) component. Responses to movement in the null direction were not consistently affected by ablation of starburst cells on either side of the directionally selective cell. (The small changes evident in these examples were inconsistent from cell to cell, see **c** and **d**.) In many experiments, cells were ablated as near to the dendrites on the null side as shown in **a**, with no effect on the ganglion cell's response. Ablation of fewer cells on the preferred side (**b**) did have an effect. Ablating cells on both sides had effects similar to the preferred side alone, again indicating a lack of effect of ablating cells on the null side. **c, d**, The integrated responses of the population of cells studied. **c**, Index of directionality (preferred-null)/preferred. **d**, Absolute responses. Responses are expressed as the total spikes per trial, with a trial consisting of one traverse of the receptive field by the stimulus. Responses to both leading edge and trailing edge are included. Ten trials were averaged for each cell, to create a value for that cell. Data shown are mean $\pm$ s.e.m. of those values for the different cells. Different groups of neurons were studied with ablation on the preferred, null or both sides of their receptive fields. The s.e.m.s are larger in **d** than **c** owing to substantial differences in the absolute rates of firing from cell to cell under both normal and experimental conditions. For individual cells, the results were quite consistent; for example, all 12 cells tested after ablation of starburst cells on the preferred side showed a depression of response.

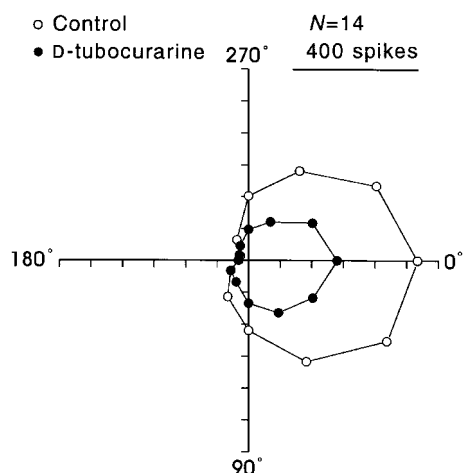


Figure 4 Effects of 50 μ M D-tubocurarine on responses of directionally selective ganglion cells to 12 directions of motion. The data are expressed as a polar plot and show mean responses for 14 cells. The directional preferences were normalized around a preferred direction of 0 deg. The distance from the origin to a data point shows the strength of the response at that angle. At this concentration, D-tubocurarine completely blocks the ganglion cell's response to large test doses of acetylcholine, carbachol or physostigmine⁷. D-tubocurarine depressed the response in many directions of stimulus movement, suggesting that cholinergic excitation from starburst cells, which are the only cholinergic neurons in the rabbit's retina^{7,29}, is not narrowly restricted to the preferred direction.

experiment shows that starburst cells on all sides of the ganglion cell's receptive field are connected (through a cholinergic synapse) to the directionally selective ganglion cells. For the starburst cells to be the source of the null-direction inhibition, the starburst cells would have to be connected on all sides and make cholinergic synapses on all sides, but make GABAergic synapses only on the null side. This is not impossible, but would require a remarkable developmental specificity. Taking all the evidence together, we conclude that the starburst cells provide neither asymmetric excitation nor asymmetric inhibition to the directionally selective ganglion cell.

We suggest that starburst cells serve as generalized, non-directional enhancers of motion sensitivity in retinal ganglion cells. The displacement of their synaptic inputs and outputs (Fig. 1) creates a signal that travels laterally across the retina. Because that signal is excitatory (Fig. 4), it should potentiate the response to moving stimuli of ganglion cells that receive it. The need for a continuous response to smooth motions would rationalize the starburst cells' densely overlapping coverage of the retina. It is sometimes forgotten that starburst amacrine cells affect many types of retinal ganglion cells, not just those that are directionally selective^{21,23,24}. The starburst cells would create a motion-sensitive excitatory substrate that is shaped by the inhibitory action of other neurons to give ganglion cells their final stimulus selectivities. In the special case of the directionally selective cell, null-direction inhibition of the ganglion cell (or of its input from bipolar cells) would come from another amacrine cell^{8,18}. Because most amacrine cells have yet to be systematically studied²⁵, there are many candidates for the direction-discriminating amacrine cell. □

Methods

Recording. Extracellular recording and intracellular dye injection were as described¹². Briefly, rabbit retinas were labelled by intraocular injection of 5 μ g DAPI about 24 h before the experiment. The retina was isolated from the pigment epithelium and attached to a coverslip treated with Cell-Tak and superfused with Ames medium²⁶ at 36 °C on a microscope stage. Tungsten-in-glass electrodes were used to record extracellular action potentials. Visual stimuli were generated on a Tektronix 608 monitor by a computer-controlled interface (Innisfree). They were reflected by a substage mirror through an Olympus $\times 20$ objective which replaced the microscope's condenser. A flashing square 100 $\times$ 100 μ m was used to map the overall receptive field. Directional

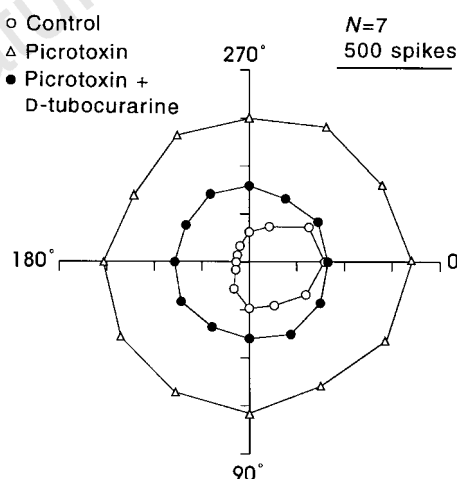


Figure 5 Effects of a cholinergic receptor antagonist in a retina treated with picrotoxin (50 μ M). This polar plot shows the means for 7 cells different from those shown in Fig. 4. As has been extensively documented^{8,22}, picrotoxin blocks null-direction inhibition, so the ganglion cell then responds equally to movements in all directions. When D-tubocurarine was applied in the presence of picrotoxin, the response to all directions of movement was depressed. The magnitude of the depression was equal in all directions, indicating that cholinergic input from starburst cells potentiates responses to all directions of movement.

responses were studied using a bar 100 $\times$ 500 μ m moving at 750 μ m s⁻¹ for a 1,500- μ m traverse in a direction parallel to its long axis. (The elongated stimulus allowed clear identification of responses to its leading or trailing edge.) Direction tuning was measured for every cell at 30-deg intervals.

Laser irradiation. Directionally selective cells had medium to large, slightly oval, somas as viewed by their weak DAPI staining. The visual cue increased the probability of recording from a directionally selective cell. Starburst amacrine cells were easy to identify in DAPI-labelled retinas, being brightly labelled with small, round cell bodies and forming a regular mosaic¹¹. To ablate the cells, a nitrogen laser, with an output at 337 nm tuned to 365 nm using a dye module (Laser Science), was used. The nominal output of the laser was 25 μ J, for 3-ns pulses at 20 Hz. The laser beam was fed into the microscope's epi-illumination port and focused by a Zeiss $\times 40$ water-immersion objective, NA 0.75, to form a 2- μ m spot on the targeted cell. During development of the technique and for the first 17 experiments, ethidium homodimer, a marker for dead cells, was added to the medium at 10 μ M. These experiments included directionally selective cells studied under all experimental conditions. In 13 subsequent retinas, we monitored the positions of the targeted starburst cells from an autofluorescent change that occurs when the cell is irradiated (Fig. 2c). We are confident that these cells were permeabilized as all experiments used a laser energy more than twice that found to permeabilize the cell to ethidium. To test the intracellular continuity of neurons after laser irradiation, we made primary cultures from newborn rat retinas, using conventional techniques²⁷. After 3 d in culture, many cells possessed thin (0.5–2.0 μ m) elongated processes. Cultures were incubated for 30 min in 5 μ M calcein AM (Molecular Probes) then transferred back to control media. The somas of 30 cells were irradiated, using the same apparatus and laser parameters as for starburst cells. In every case irradiation caused a uniform depletion of calcein from the soma and all of its processes, including some more than 200 μ m in length.

Targeting cells. To target individual starburst cells for ablation, we moved the preparation under the microscope, with the recording electrode remaining in place. We typically began the ablations 300–400 μ m from the ganglion cell receptive field and gradually worked closer. We ablated the cells as close to the ganglion cell as possible without hitting its dendrites. The number of starburst cells ablated ranged from 20 to 60. Although the starburst cells overlap extensively, only an oriented subset of their dendrites is postulated by the models to be connected to any particular retinal ganglion cell. Assuming that these are restricted to ± 30 deg from the preferred direction (as shown in Fig. 1 and refs 3, 4) the known dimensions of starburst and directionally selective cells¹² can be used to estimate the amount of 'starburst deafferentation' of the ganglion cell caused by the ablations. At the edge of the receptive field nearest the ablated cells, nearly 100% of the relevant starburst cells are eliminated. The fraction declines linearly with distance across the receptive field, reaching 0% at a point slightly (25 μ m) past its centre. Because the stimulus was narrower than the receptive field and a wide band of starburst cells was ablated, the amount of deafferentation would change little under an assumption of wider afferent starburst angles. These estimates are based on the assumption that all relevant starburst cells were ablated over a 90-deg sector centred on the preferred-null axis. Although we cannot eliminate the possibility that a few starburst cells escaped in some experiments, it is certain that the ablations caused a large reduction in starburst input to the ganglion cell over the affected half of its receptive field. This was confirmed by the clear effect of ablating starburst cells on the preferred side (Fig. 3d).

Received 24 April; accepted 23 June 1997.

- Barlow, H. B. & Levick, W. R. The mechanism of directionally selective units in rabbit's retina. *J. Physiol. (Lond.)* **178**, 477–504 (1965).
- Amthor, F. R. & Grzywacz, N. M. in *Visual Motion and its Role in the Stabilization of Gaze* (eds Miles, E. A. & Wallman, J.) (Elsevier, New York, 1993).
- Famiglietti, E. V. Jr Synaptic organization of starburst amacrine cells in rabbit retina: analysis of serial thin sections by electron microscopy and graphic reconstruction. *J. Comp. Neurol.* **309**, 40–70 (1991).
- Masland, R. H. & Tauchi, M. The cholinergic amacrine cell. *Trends Neurosci.* **9**, 218–223 (1986).
- Vaney, D. I. The mosaic of amacrine cells in the mammalian retina. *Prog. Retin. Res.* **9**, 1–28 (1991).
- Borg-Graham, L. J. & Grzywacz, N. M. in *Single Neuron Computation* (eds McKenna, T., Davis, J. & Zornetzer, S. F.) 347–376 (Academic, New York, 1992).
- Masland, R. H. & Ames, A. III Responses to acetylcholine of ganglion cells in an isolated mammalian retina. *J. Neurophysiol.* **39**, 1220–1235 (1976).
- Ariel, M. & Daw, N. W. Pharmacological analysis of directionally selective rabbit retinal ganglion cells. *J. Physiol. (Lond.)* **324**, 161–185 (1982).
- Brecha, N., Johnson, D., Peichl, L. & Wässle, H. Cholinergic amacrine cells of the rabbit retina contain glutamate decarboxylase and gamma-aminobutyrate immunoreactivity. *Proc. Natl. Acad. Sci. USA* **85**, 6187–6191 (1988).

10. O'Malley, D. M., Sandell, J. H. & Masland, R. H. Co-release of acetylcholine and GABA by the starburst amacrine cells. *J. Neurosci.* **12**, 1394–1408 (1992).
11. Masland, R. H., Mills, J. W. & Hayden, S. A. Acetylcholine-synthesizing amacrine cells: identification and selective staining by using autoradiography and fluorescent markers. *Proc. R. Soc. Lond. B* **223**, 79–100 (1984).
12. Yang, G. & Masland, R. H. Receptive fields and dendritic structure of directionally selective retinal ganglion cells. *J. Neurosci.* **14**, 5267–5280 (1994).
13. Sulston, J. E. & White, J. G. Regulation and cell autonomy during postembryonic development of *Caenorhabditis elegans*. *Dev. Biol.* **78**, 577–597 (1980).
14. Bargmann, C. I. & Avery, L. Laser killing of cells in *Caenorhabditis elegans*. *Methods Cell. Biol.* **48**, 225–250 (1995).
15. Jacobs, G. A., Miller, J. P. & Murphey, R. K. Integrative mechanisms controlling directional sensitivity of an identified sensory interneuron. *J. Neurosci.* **6**, 2298–2311 (1986).
16. Grzywacz, N. M. & Amthor, F. R. Facilitation in ON–OFF directionally selective ganglion cells of the rabbit retina. *J. Neurophysiol.* **69**, 2188–2199 (1993).
17. Marchiafava, P. L. The responses of retinal ganglion cells to stationary and moving visual stimuli. *Vision Res.* **19**, 1203–1211 (1979).
18. Torre, V. & Poggio, T. A synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond.* **202**, 409–416 (1978).
19. Cohen, E. D. & Miller, R. F. Quinoxalines block the mechanism of directional selectivity in ganglion cells of the rabbit retina. *Proc. Natl Acad. Sci. USA* **92**, 1127–1131 (1995).
20. Kittila, C. A. & Massey, S. C. The pharmacology of directionally selective ganglion cells in the rabbit retina. *J. Neurophysiol.* **77**, 675–689 (1997).
21. Ariel, M. & Daw, N. W. Effects of cholinergic drugs on receptive field properties of rabbit retinal ganglion cells. *J. Physiol. (Lond.)* **324**, 135–160 (1982).
22. Caldwell, J. H., Daw, N. W. & Wyatt, H. J. Effects of picrotoxin and strychnine on rabbit retinal ganglion cells: lateral interactions for cells with more complex receptive fields. *J. Physiol. (Lond.)* **276**, 277–298 (1978).
23. Vardi, N., Masarachia, P. H. & Sterling, P. Structure of the starburst amacrine network in the cat retina and its association with alpha ganglion cells. *J. Comp. Neurol.* **288**, 601–611 (1989).
24. Schmidt, M., Humphrey, M. F. & Wässle, H. Action and localization of acetylcholine in the cat retina. *J. Neurophysiol.* **58**, 997–1015 (1987).
25. Strettoi, E. & Masland, R. H. The number of unidentified amacrine cells in the rabbit's retina. *Proc. Natl Acad. Sci. USA* **93**, 14906–14911 (1996).
26. Ames, A. I. & Nesbitt, F. B. *In vitro* retina as an experimental model of the central nervous system. *J. Neurochem.* **37**, 867–877 (1996).
27. Huba, R., Schneider, H. & Hofmann, H. D. Voltage-gated currents of putative GABAergic amacrine cells in primary cultures and in retinal slice preparations. *Brain Res.* **577**, 10–18 (1992).
28. Tsuchi, M. & Masland, R. H. The shape and arrangement of the cholinergic neurons in the rabbit retina. *Proc. R. Soc. Lond.* **223**, 101–119 (1984).
29. Masland, R. H. & Mills, J. W. Autoradiographic identification of acetylcholine in the rabbit retina. *J. Cell Biol.* **83**, 159–178 (1979).

Acknowledgements. We thank M. Chiang for help in developing the laser system, and S. Wallenstein for preparing cultured cells.

Correspondence and requests for materials should be addressed to R.H.M. (e-mail: masland@helix.mgh.harvard.edu).

Cellular mechanism for anti-analgesic action of agonists of the κ -opioid receptor

Z. Z. Pan, S. A. Tershner & H. L. Fields

Departments of Neurology and Physiology, and the W. M. Keck Center for Integrative Neuroscience, Box 0114, University of California, San Francisco, California 94143, USA

The analgesic effect of clinically used exogenous opioids, such as morphine, is mediated primarily through μ -opioid receptors^{1–3}, but the function of the κ -receptor in opioid analgesia is unclear. Although κ -receptor agonists can produce analgesia^{4,5}, behavioural studies indicate that κ agonists applied intravenously or locally into the spinal cord antagonize morphine analgesia (see refs 4, 6 for reviews). As morphine, a primary μ agonist¹, also binds to κ -receptors⁷ and the analgesic effectiveness of morphine decreases with repeated use (tolerance), it is important to understand the mechanism for the functional interaction between κ - and μ -opioid receptors in the central nervous system. Here we present *in vitro* electrophysiological and *in vivo* behavioural evidence that activation of the κ -receptor specifically antagonizes μ -receptor-mediated analgesia. We show that in slice preparations of a rat brainstem nucleus, which is critical for the action of opioids in controlling pain, functional κ - and μ -receptors are each localized on physiologically different types of neuron. Acti-

vation of the κ -receptor hyperpolarizes neurons that are activated indirectly by the μ -receptor. In rats, κ -receptor activation in this brainstem nucleus significantly attenuates local μ -receptor-mediated analgesia. Our findings suggest a new cellular mechanism for the potentially ubiquitous opposing interaction between μ - and κ -opioid receptors and may help in the design of treatments for pain.

Using brain slice preparations (*in vitro*) and lightly anaesthetized rats (*in vivo*), we investigated the functional relationship between κ - and μ -receptors in the medullary nucleus raphe magnus (NRM) which plays a critical role in the supraspinal component of opioid analgesia^{8,9}.

In the medullary slice preparation, we used whole-cell patch clamp recordings from NRM neurons visualized with an infrared microscope to examine the cellular mechanisms of action of κ -opioid receptor agonists. U69593, a selective κ -receptor agonist, produced an outward current (hyperpolarization) in a subpopulation of NRM neurons (51/150, 34%) under voltage clamp at a holding potential of -60 mV (Table 1). Dynorphin A(1–17), a probable endogenous ligand for the κ -receptor¹⁰, induced a similar outward current in these cells (Table 1 and Fig. 1a, top). Both the dynorphin- and the U69593-induced currents were completely blocked by norbinaltorphimine (norBNI), a selective κ -receptor antagonist (Table 1 and Fig. 1a, bottom) and by naloxone ($1 \mu\text{M}$; $n = 3$ for both agonists). These observations demonstrate that the outward current is a κ -opioid receptor-mediated effect. The effect doses for dynorphin ranged from <3 nM to 100 nM (maximum), being half-maximal (EC_{50}) at 7.4 nM (Fig. 1b), consistent with a dynorphin action on the κ -receptor⁷. In voltage clamp, the U69593-induced outward current reversed its polarity at an average of -96.4 ± 1.3 mV in normal extracellular potassium concentration ($[\text{K}^+]_o$, 2.5 mM; $n = 43$; Fig. 1c, d). Higher $[\text{K}^+]_o$ (6.5 mM and 10.5 mM) shifted the reversal potential to 76.1 ± 1.9 mV ($n = 11$) and 59.2 ± 2.2 mV ($n = 7$), respectively (Fig. 1d). The shift in the reversal potential with increasing $[\text{K}^+]_o$ is in accordance with the Nernst equation (slope = -57.1). The κ -agonist-induced current had the property of barium-sensitive (not shown) inward rectification, that is, higher conductance in the inward direction measured between -100 and -120 mV, than in the outward direction between -60 and -80 mV (5.5 ± 0.4 nS and 3.9 ± 0.2 nS, respectively; $P < 0.0001$, paired *t*-test, $n = 43$; Fig. 1d). These results indicate that κ -receptor activation inhibits a subpopulation of NRM neurons by increasing an inwardly rectifying potassium conductance. This mechanism is similar to that found in spinal trigeminal neurons, where the coupling of κ -receptors to potassium channels was first reported¹¹.

We previously demonstrated that activation of the μ -receptor selectively hyperpolarizes one type of NRM neuron (secondary cell) but has no direct postsynaptic effect on the other major type (primary cell)¹². To determine which type of neuron was inhibited by κ agonists, we sequentially applied U69593 (300 nM) and the μ/δ agonist met-enkephalin (10 μM) to the same cells. (No δ -receptor-mediated postsynaptic effect was found in NRM neurons in the previous study¹² or in this study ($n = 8$)). Figure 2 shows an example from each of the two types of cell. In a primary cell under voltage clamp (Fig. 2a), U69593 produced an outward current, but met-enkephalin had no effect on the cell. In contrast, a secondary cell was hyperpolarized by met-enkephalin, but was not affected by the κ agonist (Fig. 2b). As summarized in the Table 2, none of the neurons ($n = 14$) that were hyperpolarized by U69593 were affected by met-enkephalin; of those cells that were inhibited by met-enkephalin ($n = 21$), the majority ($n = 20$) showed no response to U69593. These findings demonstrate that functional κ - and μ -receptors segregated into two physiologically distinct types of NRM neuron. We have previously shown that primary cells have a GABAergic synaptic input that is inhibited by μ -receptor activation¹². Our results indicate that κ -receptor-mediated

inhibition can counteract μ -receptor-induced disinhibition (excitation) of primary neurons in the NRM.

In rats, morphine or other μ -receptor agonists such as DAMGO, given either systemically or locally into the midbrain periaqueductal grey (PAG) or into the NRM, consistently produce analgesia and inhibit one type of NRM cell while disinhibiting another type^{8,9,13}. The disinhibited type is essential for supraspinal opioid analgesia through direct projection to the spinal cord dorsal horn^{8,9}. If, as we found, κ -receptor agonists exclusively inhibit the NRM cells that are activated (disinhibited) by μ -receptor agonists, κ -receptor activation in the NRM should antagonize μ -receptor-mediated analgesia. We tested this hypothesis in lightly anaesthetized rats *in vivo*.

Microinjection of the μ -receptor agonist DAMGO ($0.05 \mu\text{g}$ $0.25 \mu\text{l}^{-1}$) into the PAG induced a strong analgesic effect, as

measured by a time-dependent increase in the tail-flick latency ($n = 5$; Fig. 3). As shown in Fig. 3, the κ agonist U69593 ($0.178 \mu\text{g} \mu\text{l}^{-1}$), microinjected into the NRM just before DAMGO injection in the PAG, significantly attenuated DAMGO-induced analgesia ($n = 5$); however, U69593 microinjected in the NRM before saline injection in the PAG did not alter the pain threshold ($n = 5$; Fig. 3). The antagonizing effect of U69593 was completely prevented by microinjection of the κ antagonist norBNI ($0.367 \text{ ng} \mu\text{l}^{-1}$) into the NRM 30 min before U69593 injection ($n = 5$; Fig. 3). These drug effects were statistically confirmed by an ANOVA for repeated measures (a group by trial interaction; $P < 0.001$) and by *post hoc* contrasts (Fig. 3). Analgesia induced through the PAG is mediated by an endogenous opioid acting at the μ -receptor in the NRM as it is attenuated by the μ -receptor antagonists β -funaltrexamine or CTOP applied into the NRM^{14,15}.

Figure 1 Agonists of the κ -receptor hyperpolarize a subpopulation of NRM neurons by increasing potassium conductance. **a**, In voltage clamp (holding potential, -60 mV), dynorphin (100 nM) induced an outward current (top) that was completely blocked by norbinaltorphimine (100 nM ; bottom). **b**, Dose-response curve for the dynorphin effect. Data points (mean $\pm$ s.e., $n = 8$) were fitted with the logistic equation. **c**, Current traces in control and in U69593 (300 nM)-treated cell during voltage steps from the holding potential (-60 mV) to potentials as indicated. **d**, A current-voltage plot in control and in U69593-treated cell at two extracellular K^+ concentrations.

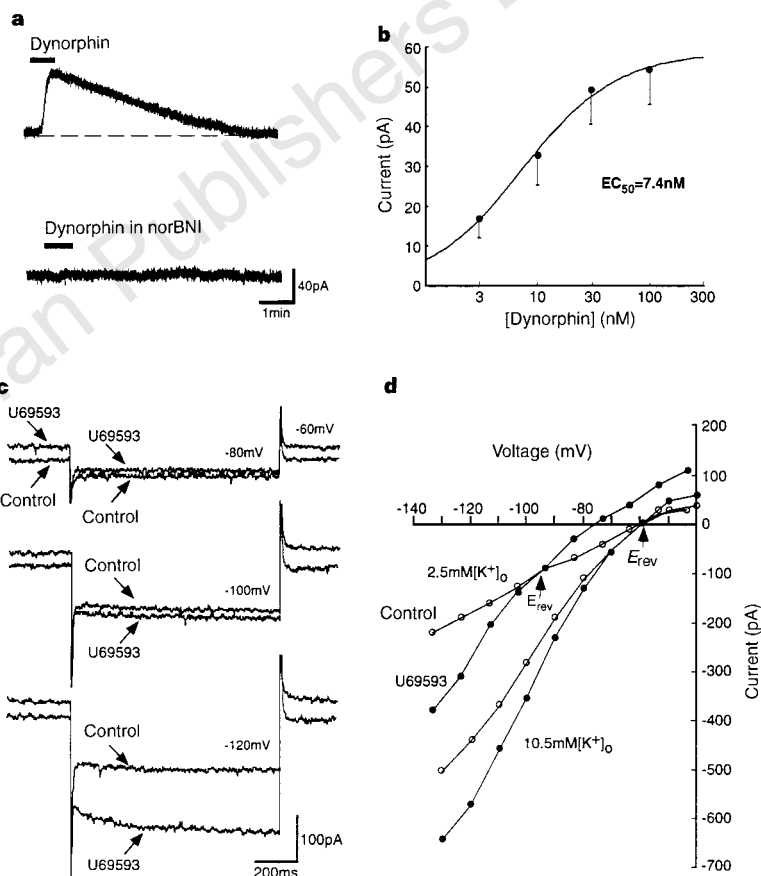


Table 1 Inhibition (outward current) induced by κ -receptor agonists

	Control	Norbinaltorphimine (100 nM)		<i>n</i>
		Before	After	
U69593 (300 nM)	$29 \pm 3 \text{ pA}$ ($n = 51$)	$39 \pm 6 \text{ pA}$	$1.1 \pm 0.9 \text{ pA}$	7
Dynorphin (100 nM)	$59 \pm 9 \text{ pA}$ ($n = 9$)	$59 \pm 12 \text{ pA}$	$0.2 \pm 1.3 \text{ pA}$	5

Table 2 Effects of κ and μ agonists on two types of NRM neurons

	U69593		Met-enkephalin	
	Inhibition	No effect	Inhibition	No effect
Primary cell ($n = 14$)	14 ($+37 \pm 8 \text{ pA}$)	(NA)	0/14	14/14 ($+0.5 \pm 1.1 \text{ pA}$)
Secondary cell ($n = 21$)	1/21 ($-1.3 \pm 1.0 \text{ pA}$)	20/21	21 ($+24 \pm 3 \text{ pA}$)	(NA)

Outward current is indicated by a plus sign; inward current by a minus sign. Only those cells in which effects of both agonists were sequentially examined are included. NA, not applicable.

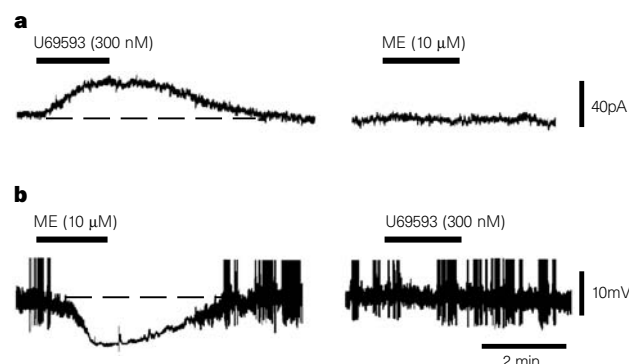


Figure 2 Activation of κ - and μ -receptors inhibit two distinct types of NRM cell respectively. **a**, In a primary cell under voltage clamp, U69593 produced an outward current, but met-enkephalin (ME) had no effect. **b**, A secondary cell was hyperpolarized by ME in current clamp, but was not affected by U69593. Spontaneous action potentials are truncated.

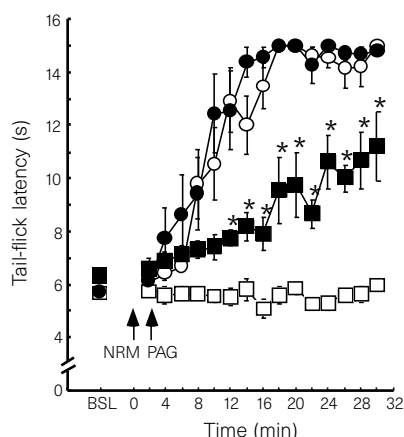


Figure 3 Activation of κ -receptor in the NRM antagonizes the analgesia induced by DAMGO microinjected into the periaqueductal grey (PAG). Tail-flick latencies (mean $\pm$ s.e.) were recorded from each ($n = 5$ rats) of the four groups (NRM injection/PAG injection): $\circ$, saline/DAMGO; $\blacksquare$, U69593/DAMGO; $\bullet$, norBNI + U69593/DAMGO; $\square$, U69593/saline. Asterisks denote data points in the U69593/DAMGO group that are statistically different from those in both the saline/DAMGO group and the norBNI + U69593/DAMGO group ($P < 0.05$, *post hoc* contrasts). BSL, means of six baseline trials in each group before drug treatment.

Thus, our findings indicate that in the NRM κ -receptor activation antagonizes μ -receptor-mediated analgesia. They confirm the behavioural significance of the opposing interaction between κ - and μ -receptors that we found in the same nucleus *in vitro*. Figure 4 shows a likely mechanism for this interaction in the modulation of pain.

The antagonism by κ -receptor ligands of μ -receptor-mediated actions has also been reported in opioid tolerance studies in which systemic application of κ -receptor agonists blocks the development of morphine tolerance^{4,6,16}. The opposing effect on μ of the κ -receptor extends to opioid actions on emotion, perception and drug reinforcement both in animals and in humans. Thus, whereas morphine and other μ agonists produce euphoria and place preference^{17,18}, κ agonists cause dysphoria and aversion¹⁹; morphine-like opioids possess the property of being rewarding²⁰ and increase dopamine release^{21,22}, whereas κ -receptor agonists block morphine-induced rewarding effects^{23,24} and decrease dopamine release in the mesolimbic system^{21,22}, a brain region closely associated with reward behaviour²⁰. In view of the broad presence of this opposing relationship between μ - and κ -opioid receptors in the brain, the neural mechanism of μ/κ antagonism identified here may be common to endogenous opioid actions throughout the central nervous system. In addition to changes in function of the μ -receptor itself and its associated second-messenger systems^{25–28}, an imbalance between μ - and κ -receptors may contribute significantly to the adaptive changes resulting from long-term use of opioids. A more comprehensive understanding of endogenous opioid receptor systems should improve treatment of opioid-related clinical problems such as analgesic tolerance and opioid addiction. $\square$

Methods

Whole-cell recordings. Brainstem slices (200 μ m thick) were cut from a male neonatal (9–14 days old) rat brain in a vibrotome (series 1000, Technical Products International) in cold physiological saline at 4°C. A single slice was submerged in a shallow recording chamber through which flowed preheated (35°C) physiological saline (in mM: NaCl, 126; KCl, 2.5; NaH₂PO₄, 1.2; MgCl₂, 1.2; CaCl₂, 2.4; glucose, 11; NaHCO₃, 25, pH 7.3, saturated with 95% O₂ and 5% CO₂). Neurons in the nucleus raphe magnus (NRM, a triangular area in the midline above the pyramids) were visualized using infrared light and Nomarski optics (Carl Zeiss). Whole-cell patch clamp recordings were made with a glass pipette (resistance, 3–5 M Ω) filled with a solution containing (in mM):

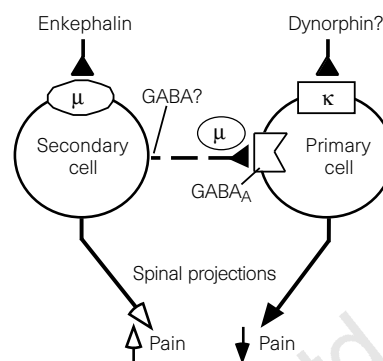


Figure 4 Model for the opposing interaction between agonists of μ - and κ -opioid receptor in the NRM. Primary cells, which inhibit spinal pain transmission, are excited (disinhibited) by μ -receptor-mediated presynaptic inhibition of a GABAergic input. Secondary cells, which facilitate pain transmission⁸, are inhibited by μ -receptor activation. Both μ actions produce analgesia. Agonists of κ -receptors selectively inhibit primary cells, thus reversing the disinhibition and analgesia produced by μ agonists. Secondary cells receive an enkephalinergic input and we propose that primary cells receive a dynorphin input and a GABAergic input from a subset of secondary cells.

potassium gluconate, 126; NaCl, 10; MgCl₂, 1; EGTA, 11; HEPES, 10; ATP, 2; GTP, 0.25; pH adjusted to 7.3 with KOH. A seal resistance of ≥ 5 G Ω and an access resistance of ≤ 15 M Ω were considered acceptable. Series resistance was optimally compensated by at least 70% with an Axopatch 1D amplifier (Axon Instruments). All drugs were applied through the perfusing solution. Neonatal rats were used for better visualization of neurons with the microscope. In slices from these young rats, the properties of NRM neurons were indistinguishable from those of adult rats¹²; two similar types of cell (primary and secondary cells) were identified. Met-enkephalin hyperpolarized only the secondary cell (see text); μ agonists did not affect the primary cell postsynaptically, but inhibited the bicuculline-sensitive GABA synaptic potential in primary cells ($n = 5$). The κ agonist actions were observed in NRM neurons from adult rats: κ agonists hyperpolarized the primary cells (-6.2 ± 1.1 mV; $n = 5$) and did not affect the secondary cells ($n = 4$).

Microinjections. Male Sprague-Dawley rats (325–450 g) were maintained lightly anaesthetized in a stereotaxic apparatus with a constant i.v. infusion of methohexital (10 mg ml⁻¹ at 0.8 ml h⁻¹). Two 26-gauge guide cannulae were aimed at the ventrolateral PAG (AP: -7.8; L: 0.8; V: 6.0) and NRM area (AP: -11.0; L: 0.5; V: 10.7) (ref. 29; AP, anterior (+) or posterior (-) to Bregma; L, lateral to midline; V, ventral to the surface of skull; in mm). Tail-flick latency to a radiant heat stimulus was measured every 2 min. The heat intensity was set to elicit stable baseline latencies between 5.0 and 6.5 s, with a cut-off time of 15 s. After six baseline trials, drugs were delivered through a 33-gauge cannula with an infusion pump at a rate of 0.5 μ l min⁻¹. All cannula placements for both the PAG and the NRM were histologically verified afterwards and only those located within the designated area were included in the results.

Materials. All drugs used were purchased from Research Biochemicals International and from Sigma.

Received 16 June; accepted 21 July 1997.

- Matthes, H. W. D. *et al.* Loss of morphine-induced analgesia, reward effect and withdrawal symptoms in mice lacking the μ -opioid-receptor gene. *Nature* **383**, 819–823 (1996).
- Reisine, T. & Pasternak, G. in *Goodman and Gilman's Pharmacological Basis of Therapeutics* (eds Hardman, J. G. & Limbird, L. E.) 521–555 (McGraw-Hill, New York, 1996).
- Zadina, J. E., Hackler, L., Ge, L.-J. & Kastin, A. J. A potent and selective endogenous agonist for the μ -opioid receptor. *Nature* **386**, 499–502 (1997).
- Bhargava, H. N. Diversity of agents that modify opioid tolerance, physical dependence, abstinence syndrome and self-administrative behavior. *Pharmacol. Rev.* **46**, 293–324 (1994).
- Millan, M. J. κ -Opioid receptors and analgesia. *Trends Pharmacol. Sci.* **11**, 70–76 (1990).
- Rothman, R. B. A review of the role of anti-opioid peptides in morphine tolerance and dependence. *Synapse* **12**, 129–138 (1992).
- Goldstein, A. & Naidu, A. Multiple opioid receptors: ligand selectivity profiles and binding site signatures. *Mol. Pharmacol.* **36**, 265–272 (1989).
- Basbaum, A. I. & Fields, H. L. Endogenous pain control systems: Brainstem spinal pathways and endorphin circuitry. *Annu. Rev. Neurosci.* **7**, 309–338 (1984).
- Fields, H. L., Heinricher, M. M. & Mason, P. Neurotransmitters in nociceptive modulatory circuits.

- Annu. Rev. Neurosci.* **14**, 219–245 (1991).
10. Chavkin, C., James, I. F. & Goldstein, A. Dynorphin is a specific endogenous ligand of the κ -opioid receptor. *Science* **215**, 413–415 (1982).
 11. Grudt, T. J. & Williams, J. T. κ -Opioid receptors also increase potassium conductance. *Proc. Natl Acad. Sci. USA* **90**, 11429–11432 (1993).
 12. Pan, Z., Williams, J. T. & Osborne, P. Opioid acts on single raphe magnus neurons from rat and guinea pig *in vitro*. *J. Physiol. (Lond.)* **427**, 519–532 (1990).
 13. Fields, H. L., Vanegas, H., Hentall, I. D. & Zorman, G. Evidence that disinhibition of brain stem neurons contributes to morphine analgesia. *Nature* **306**, 684–686 (1983).
 14. Kiehl, J. M., Rossi, G. C. & Bodnar, R. J. Medullary μ - and δ -opioid receptors modulate mesencephalic morphine analgesia in rats. *Brain Res.* **624**, 151–161 (1993).
 15. Roychowdhury, S. M. & Fields, H. L. Endogenous opioids acting at a medullary μ -opioid receptor contribute to the behavioral antinociception produced by GABA antagonism in the midbrain periaqueductal gray. *Neuroscience* **74**, 863–872 (1996).
 16. Hooke, L. P., He, L. & Lee, N. M. Dynorphin A modulates acute and chronic opioid effects. *J. Pharmac. Exp. Ther.* **273**, 292–297 (1995).
 17. Martin, W. R. Pharmacology of opioids. *Pharmacol. Rev.* **35**, 283–323 (1983).
 18. Koob, G. F. Drugs of abuse: anatomy, pharmacology and function of reward pathways. *Trends Pharmacol. Sci.* **13**, 177–184 (1992).
 19. Pfeiffer, A., Brantl, V., Herz, A. & Emrich, H. M. Psychotomimesis mediated by κ opiate receptors. *Science* **233**, 774–776 (1986).
 20. Di Chiara, G. & North, R. A. Neurobiology of opiate abuse. *Trends Pharmacol. Sci.* **13**, 185–193 (1992).
 21. Spanagel, R., Herz, A. & Shippenberg, T. S. Opposing tonically active endogenous opioid systems modulate the mesolimbic dopaminergic pathway. *Proc. Natl Acad. Sci. USA* **89**, 2046–2050 (1992).
 22. Di Chiara, G. & Imperato, A. Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats. *Proc. Natl Acad. Sci. USA* **85**, 5274–5278 (1988).
 23. Funada, M., Suzuki, T., Narita, M., Misawa, M. & Nagase, H. Blockade of morphine reward through the activation of κ -opioid receptors in mice. *Neuropharmacology* **32**, 1315–1323 (1993).
 24. Bolanos, C. A., Garmsen, G. M., Clair, M. A. & McDougall, S. A. Effects of the kappa-opioid receptor agonist U-50,488 on morphine-induced place preference conditioning in the developing rat. *Eur. J. Pharmacol.* **317**, 1–8 (1996).
 25. Chang, K. J., Eckel, R. W. & Blanchard, S. G. Opioid peptides induce reduction of enkephalin receptors in cultured neuroblastoma cells. *Nature* **296**, 446–448 (1982).
 26. Loh, H. & Smith, A. Molecular characterization of opioid receptors. *Annu. Rev. Pharmacol. Toxicol.* **30**, 123–147 (1990).
 27. Sharma, S. K., Klee, W. & Nirenberg, M. Dual regulation of adenylate cyclase accounts for narcotic dependence and tolerance. *Proc. Natl Acad. Sci. USA* **72**, 3092–3096 (1975).
 28. Avidor-Reiss, T. et al. Adenylyl cyclase supersensitization in μ -opioid receptor-transfected Chinese hamster ovary cells following chronic opioid treatment. *J. Biol. Chem.* **270**, 29732–29738 (1995).
 29. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* 2nd edn (Academic, Sydney, 1986).

Acknowledgements. We thank R. Nicoll for reviewing the manuscript. This work was supported by the National Institute of Drug Abuse.

Correspondence and requests for materials should be addressed to Z.Z.P. (e-mail: zzp@itsa.ucsf.edu).

Sites of alcohol and volatile anaesthetic action on GABA_A and glycine receptors

S. John Mihic*[‡], Qing Ye[†], Marilee J. Wick*[‡], Vladimir V. Koltchine[§], Matthew D. Krasowski^{||}, Suzanne E. Finn[†], Maria Paola Mascia*, C. Fernando Valenzuela*, Kirsten K. Hanson[†], Eric P. Greenblatt[✱], R. Adron Harris*[‡] & Neil L. Harrison[†][§]

* Department of Pharmacology and ‡ Alcohol Research Center, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

Departments of † Anesthesia and Critical Care, § Pharmacological and Physiological Sciences, and || Neurobiology, The University of Chicago, Chicago, Illinois 60637, USA

‡ Denver Veteran Administration Medical Center, Alcoholism Research Center, Denver, Colorado 80220, USA

✱ Department of Anesthesia, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Volatile anaesthetics have historically been considered to act in a nonspecific manner on the central nervous system^{1,2}. More recent studies, however, have revealed that the receptors for inhibitory neurotransmitters such as γ -aminobutyric acid (GABA) and glycine are sensitive to clinically relevant concentrations of inhaled anaesthetics³. The function of GABA_A and glycine recep-

tors is enhanced by a number of anaesthetics^{4–9} and alcohols^{10–12}, whereas activity of the related¹³ GABA ρ 1 receptor is reduced¹⁴. We have used this difference in pharmacology to investigate the molecular basis for modulation of these receptors by anaesthetics and alcohols. By using chimaeric receptor constructs, we have identified a region of 45 amino-acid residues that is both necessary and sufficient for the enhancement of receptor function. Within this region, two specific amino-acid residues in transmembrane domains 2 and 3 are critical for allosteric modulation of both GABA_A and glycine receptors by alcohols and two volatile anaesthetics. These observations support the idea that anaesthetics exert a specific effect on these ion-channel proteins, and allow for the future testing of specific hypotheses of the action of anaesthetics.

We constructed a chimaeric receptor (which we call C1) that was composed of the glycine-receptor α 1 sequence from the amino terminus of the receptor to the junction site in the third of the four transmembrane domains (TM3) believed to be present in each of the receptor subunits; from the junction site to the carboxy terminus, it contained GABA-receptor ρ 1 sequence (Fig. 1). When this chimaera was expressed in *Xenopus* oocytes it was activated by glycine but not by GABA. Ethanol and the volatile anaesthetic enflurane increased the action of glycine on C1, suggesting that they act on the N-terminal side of the junction site; that is, chimaera C1 behaved more like a glycine α 1-subunit receptor than a ρ 1 receptor with regard to anaesthetic and ethanol modulation. Chimaera C2 was the converse of C1, consisting of GABA-receptor ρ 1 sequence from the N terminus of the subunit to the junction site in TM3, with the remainder being Gly-receptor α 1 sequence (Fig. 1). As expected, chimaera C2 was activated by GABA but not by glycine. Ethanol and enflurane inhibited currents evoked by GABA in the C2 chimaera, confirming the conclusion from C1 that these agents act on the N-terminal side of the junction site in TM3. The third chimaera, C3, was composed of GABA-receptor ρ 1 sequence to the junction site in TM1, and Gly-receptor α 1 sequence thereafter (Fig. 1). The C3 chimaera was activated by GABA but not glycine. Ethanol and enflurane potentiated GABA-activated currents in C3, but not wild-type ρ 1. We concluded that ethanol and enflurane act on the C-terminal side of the junction site in TM1 in this chimaera. Data obtained from these three chimaeras suggest that modulation by ethanol and enflurane of GABA ρ 1 and glycine α 1 receptors was dependent on one or more of the 63 amino-acid residues lying between the middle of TM1 and the C-terminal end of TM3 (Fig. 1).

The importance of TM2 and TM3 was later confirmed with chimaeras C4 and C5, in which the junction site was in the short loop between TM1 and TM2. C4, which contains Gly-receptor α 1 sequence on the N-terminal side of the junction site, displayed inhibition of receptor function by ethanol and enflurane, whereas the converse chimaera, C5, exhibited enhancement (Fig. 1). To investigate further the involvement of TM2 and TM3 domains of the Gly-receptor α 1 subunit in the alcohol and volatile anaesthetic enhancement of receptor function, chimaera C6 was constructed. C6 was a ρ 1 subunit containing only the TM2 and TM3 of the Gly-receptor α 1 subunit (Fig. 1). Because chimaera C6 exhibited potentiation by both ethanol and enflurane, we conclude that a segment of 45 amino-acid residues within the TM2 and TM3 domains of Gly-receptor α 1 is both necessary and sufficient to confer enhancement by alcohol and enflurane of glycine-receptor function.

Anaesthetics have been shown to enhance the function of homomeric GABA-receptor β 1 subunits¹⁵ and glycine receptors¹², suggesting that they have some common sites of anaesthetic action. We compared the amino-acid sequences of TM1–TM3 in the Gly-receptor α 1, GABA_A-receptor α 1, α 2, β 1, β 3 and GABA-receptor ρ 1 subunits. Our reasoning was that amino-acid residues that are the same in the GABA-receptor β 1 and Gly-receptor α 1 subunits, but different in GABA-receptor ρ 1, are most likely to be responsible for the lack of anaesthetic potentiation seen in the ρ 1 receptors (Fig. 2).

Present address: Department of Physiology and Pharmacology, Bowman Gray School of Medicine of Wake Forest University, Winston-Salem, North Carolina 27157, USA

Figure 1 Effects of ethanol and enflurane on wild-type glycine $\alpha 1$, GABA $\rho 1$ and chimaeric receptors. Ethanol and enflurane potentiated the effects of glycine on Gly-receptor (Gly-R) $\alpha 1$ (white bars) and inhibited the effect of GABA-receptor $\rho 1$ (black bars). The numbers on the right represent the percentage change of the control agonist response induced by ethanol (200 mM) and by enflurane (1 mM) in each of the receptors tested. Linear schematic representations of each chimaera are shown; numbers indicate transmembrane domains. Interfaces between white and black bars indicate chimaera junction sites, also illustrated in the topological diagram below. Data for C6 were obtained in HEK 293 cells ($n = 4-6$).

Table 1: Domain Lengths (Residues) for Various Cys-loop Subunits

Subunit	Ethanol	Enflurane
$\alpha 1$	+88 $\pm$ 11	+128 $\pm$ 15
R p1	-73 $\pm$ 6	-25 $\pm$ 5
C1	+47 $\pm$ 7	+64 $\pm$ 9
C2	-80 $\pm$ 4	-34 $\pm$ 6
C3	+35 $\pm$ 7	+55 $\pm$ 8
C4	-13 $\pm$ 10	-16 $\pm$ 3
C5	+33 $\pm$ 1	+37 $\pm$ 8
C6	+48 $\pm$ 18	+106 $\pm$ 18

Figure 1: Schematic representation of the M1-M4 domain organization of the Cys-loop. The diagram shows a ribbon diagram of the Cys-loop structure with domains M1, M2, M3, and M4 labeled. Arrows indicate the positions of C3, C4, C5, C6, C1, C2, and C6 relative to the domains.

Figure 2 Sequence alignment of the TM2 and TM3 domains that are important for the action of ethanol and enflurane. An alignment of the Gly-receptor $\alpha 1$, GABA-receptor $\alpha 1$, $\alpha 2$, $\beta 1$, $\beta 3$ and $\rho 1$ polypeptide amino-acid sequences within this general region highlighted 14 amino acids (indicated by asterisks) that are identical in Gly-receptor $\alpha 1$ and GABA-receptor $\beta 1$, but different in $\rho 1$. Two other amino acids (indicated by hash signs), which differ between the glycine $\alpha 1$ and $\rho 1$ subunits, were also considered to be potentially relevant for alcohol and anaesthetic effects.

				TM1			TM2		
				**	**	#	*	*	
Human	Gly	α 1	231	SLILVILSWI	SWFINMDAAP	ARVLGGITTV	LTMTTQSSGS	270	
Human	GABA	α 1	234	CIMTVILSQV	SWFLNRESVP	ARTVFGVTVT	LTMTTLLSIA	273	
Human	GABA	α 2	234	CIMTVILSQV	SWFLNRESVP	ARTVFGVTVT	LTMTTLLSIA	273	
Human	GABA	β 1	229	SLITILISWV	SWFINYDASA	ARVALGITTV	LTMTTITSTH	268	
Human	GABA	β 3	229	SILITILISWV	SWFINYDASA	ARVALGITTV	LTMTTITNTH	268	
Human	GABA	ρ 1	271	ATLMVMLSWV	SWFIDRRAPV	ARVPLGGITTV	LTMTSTIITG	310	
TM3									
				*	*	*	*	*	
Human	Gly	α 1	271	RASLPKVSIVY	KAIDINWMAVC	LLFVFSALLE	300		
Human	GABA	α 1	274	RNSLPKWAYA	TAMDWFIACV	YAFVFSALIE	303		
Human	GABA	α 2	274	RNSLPKWAYA	TAMDWFIACV	YAFVFSALIE	303		
Human	GABA	β 1	269	RETLPKPIPV	KAIDIIYLMGC	FVFVFLALLE	298		
Human	GABA	β 3	269	RETLPKPIPV	KAIDIIYLMGC	FVFVFLALLE	298		
Human	GABA	ρ 1	311	RNSPRVSIYI	KAVIDIYLWVS	FVFVFLSVLE	340		

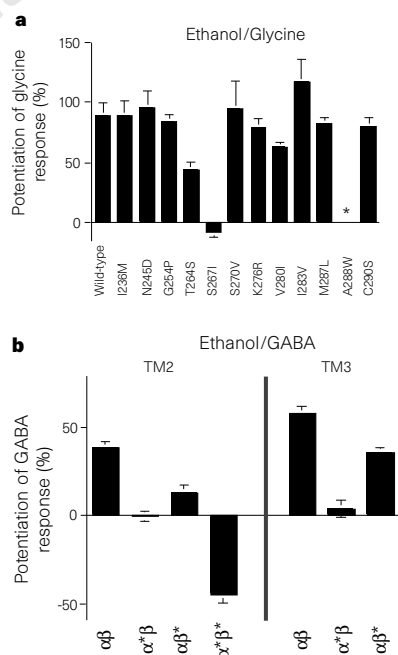


Figure 3 Identification of amino-acid residues in glycine and GABA_A receptors critical for the actions of ethanol. **a**, Mutation of 12 individual amino-acid residues of the Gly-receptor $\alpha 1$ subunit (highlighted in Fig. 2) to their $\rho 1$ counterparts identified Ser 267 as being critical for the modulatory actions of ethanol. In the wild-type receptor, 20–200 mM ethanol increased glycine-receptor currents. All mutants were tested with 200 mM ethanol. The D247R mutant failed to express; A288W (asterisk) was unusual in producing channels that were open in the absence of agonist; and R271N and L274M were extremely insensitive to glycine; as a result, none of these were tested for ethanol effects. **b**, Expression of either GABA_A-receptor $\alpha 1$ (S270I) or $\beta 1$ (S265I) TM2 mutants significantly decreased ($P < 0.001$) the enhancement of submaximal GABA responses by 200 mM ethanol. Expression of either GABA_A-receptor $\alpha 2$ (A291W) or $\beta 1$ (M286W) TM3 mutants also significantly decreased ($P < 0.001$) the enhancement of submaximal GABA responses by 200 mM ethanol. Mutant subunits are indicated by an asterisk.

of 240 ± 80 and $155 \pm 17 \mu\text{M}$, respectively ($n = 6$). The S267I mutation did not affect the reversal potential for glycine-activated currents, and the half-maximal inhibitory concentration (IC_{50}) of the Gly-receptor antagonist strychnine was $26.9 \pm 3.9 \text{ nM}$ ($n = 9$) in the wild-type Gly-receptor and $86.9 \pm 15.9 \text{ nM}$ ($n = 10$) in the S267I mutant.

Studies on anaesthetic binding sites in proteins of known structure lead us to hypothesize that an alcohol-binding pocket in glycine receptors could be formed between two or more transmembrane domains, and so we examined TM3 for residues that might also be important. The Gly-receptor $\alpha 1$ subunit has an alanine residue at position 288 in TM3, whereas the corresponding residue in GABA-R $\rho 1$ is tryptophan. Because of the non-conservative nature of this substitution, we constructed homomeric glycine $\alpha 1$ A288W mutant receptors (Fig. 3a) to test for ethanol potentiation. These receptors were tonically active in the absence of glycine, precluding tests of ethanol modulation of glycinergic function. Application of the chloride-channel blocker picrotoxin induced an outward current, consistent with closure of tonically open channels. Ethanol and enflurane, applied in the absence of glycine, were also found to inhibit the tonic activity of this mutant receptor (data not shown).

To determine whether residues homologous with Ser267 and Ala288 in Gly-receptor $\alpha 1$ were essential for the action of ethanol in GABA_A receptors, we next made mutants at the equivalent positions in GABA-receptor α and β subunits. GABA_A receptors consisting of wild-type $\alpha 1\beta 1$ subunits show modest potentiation by ethanol (Fig. 3b). Coexpression of wild-type $\beta 1$ with the $\alpha 1$ (S270I) mutant produced receptors resistant to ethanol enhancement (Fig. 3b), and similar results were obtained using $\alpha 2$ (S270I) $\beta 1$ receptors (data not shown). Coexpression of wild-type $\alpha 1$ with $\beta 1$ (S265I) subunits reduced ethanol enhancement of the resulting receptors

to a lesser degree (Fig. 3b), and similar results were obtained with $\alpha 1\beta 3$ (N265I) receptors. GABA_A receptors comprising both $\alpha 1$ (S270I) and $\beta 1$ (S265I) subunits exhibited inhibition by ethanol of the effects of GABA.

We reasoned that the Gly-receptor $\alpha 1$ (A288W) mutant exhibits unusual behaviour perhaps because it contains one additional bulky tryptophan residue in each of the five TM3 domains, and that mutant receptors containing fewer Trp residues would show normal ligand gating. The A288W Gly-receptor contains ten Trp residues in TM3, whereas the wild-type GABA $\rho 1$ receptor contains only five, perhaps explaining why the latter channel is not tonically open. We mutated the homologous residues in TM3 of the GABA-receptor α and β subunits to Trp and found that coexpression of GABA_A-receptor $\alpha 2$ (A291W) with wild-type $\beta 1$ produced a receptor with normal ligand gating; indeed, this mutant had increased sensitivity to GABA. These receptors were completely insensitive to ethanol (Fig. 3b). Coexpression of GABA_A-receptor $\beta 1$ (M286W) with wild-type $\alpha 2$ produced receptors in which ethanol still had some effects (Fig. 3b). Coexpression of the GABA_A-receptor α (A291W) and β (M286W) mutant subunits produced a receptor with unusual gating properties, similar to those of Gly-receptor $\alpha 1$ (A288W). The current associated with the resting state of this receptor was blocked by 200 mM ethanol in the absence of GABA.

Modulation by enflurane was associated with the same 45 amino-acid domain as the action of ethanol (Fig. 1). All of the original Gly-receptor mutants tested retained normal or increased sensitivity to enflurane (Fig. 4a). Although the S267I mutation rendered the glycine $\alpha 1$ receptor resistant to the potentiating effects of ethanol, it did not markedly affect potentiation by enflurane. However, substitution of Ser 267 with tyrosine, an even larger amino-acid residue than isoleucine, produced a mutant glycine receptor (S267Y) that

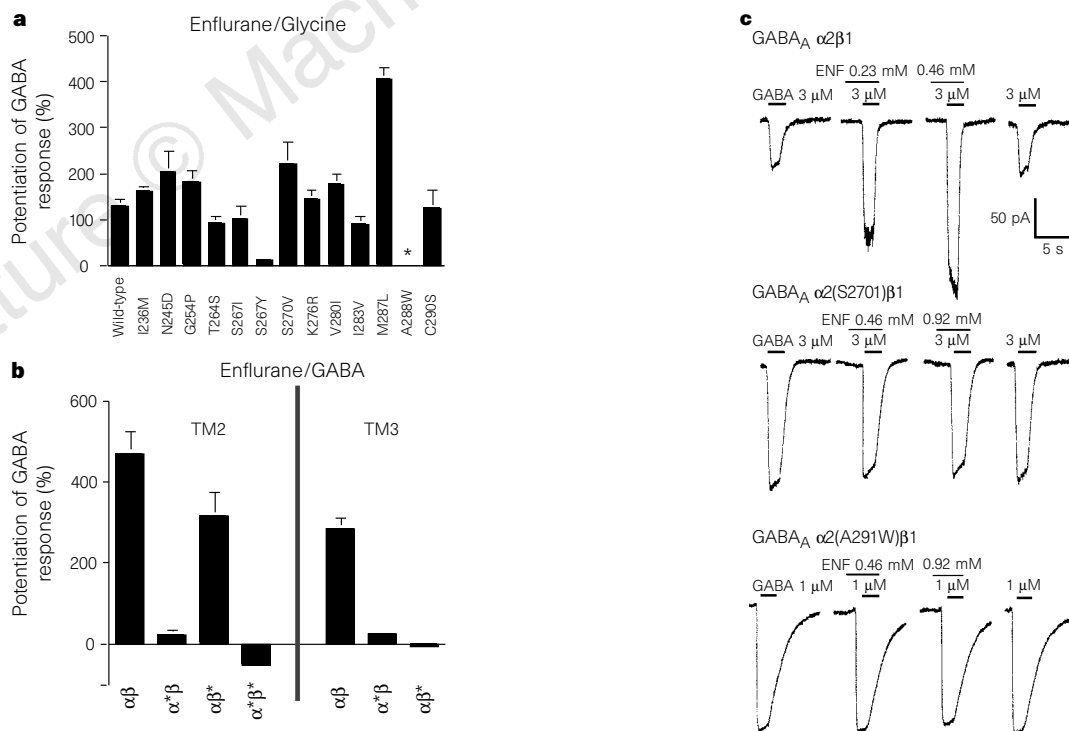


Figure 4 Amino-acid residues in glycine and GABA_A receptors critical for the action of the volatile anaesthetic enflurane. **a**, Mutation of Ser 267 to tyrosine, but not to isoleucine, abolished the effect of enflurane on Gly-receptor $\alpha 1$. All other mutants tested responded normally or supranormally to 1 mM enflurane. **b**, Expression of either GABA_A-receptor $\alpha 1$ (S270I) or $\beta 1$ (S265I) TM2 mutants significantly decreased ($P < 0.001$) the enhancement by 1 mM enflurane. Expression of either GABA_A-receptor $\alpha 1$ (A291W) or $\beta 1$ (M286W) TM3 mutants also abolished ($P < 0.001$) the enhancement of GABA responses by 1 mM enflurane,

but with no difference between the effects of mutating either the α or β mutant. Mutant subunits are depicted by an asterisk. **c**, Recordings from HEK 293 cells ($n = 4$) coexpressing wild-type GABA_A $\alpha 2\beta 1$ receptors or mutant $\alpha 2$ (S270I) $\beta 1$ or $\alpha 2$ (A291W) $\beta 1$ receptors. Wild-type $\alpha 2\beta 1$ receptors (top) show significant potentiation by 0.23 and 0.46 mM enflurane, unlike $\alpha 2$ (S270I) $\beta 1$ (middle) and $\alpha 2$ (A291W) $\beta 1$ (bottom) receptors. Horizontal bars indicate the period of drug application. An EC_{20} concentration of GABA for each receptor was used in these experiments.

was virtually insensitive to 1 mM enflurane (Fig. 4a), suggesting that Ser 267 is also important for the actions of enflurane on Gly-receptors. The M287L mutation exhibited significantly greater enhancement by enflurane than the wild-type Gly-receptor (Fig. 4a), but this was not observed for ethanol (Fig. 3a).

We then investigated the effects of mutations in TM2 and TM3 of GABA_A-receptor α and β subunits on sensitivity to enflurane. Mutation of Ser 270 in TM2 of GABA_A-receptor $\alpha 1$ or $\alpha 2$ to isoleucine completely abolished 1 mM potentiation by enflurane: the same mutation in GABA_A-receptor $\beta 1$ had a much weaker effect (Fig. 4b). Coexpression of these two TM2 mutants produced a receptor with normal GABA gating but that exhibited inhibition by enflurane (Fig. 4b). Mutation of Ala 291 in TM3 of GABA_A-receptor $\alpha 2$ or Met 286 in TM3 of GABA_A-receptor $\beta 1$ to tryptophan completely abolished enflurane potentiation (Fig. 4b, c). Similar results were obtained in each case with another volatile anaesthetic, isoflurane (data not shown).

Finally, we investigated whether these mutations in GABA_A receptors were specific to ethanol or enflurane, or whether they caused general insensitivity to multiple anaesthetic modulators. The intravenous anaesthetic propofol, like the volatile anaesthetics, enhances GABA_A receptor function⁴ in a manner that does not depend on the presence of the γ subunit in the receptor^{15,16}. The two point mutations GABA_A-receptor $\alpha 2$ (S270I) and GABA_A-receptor $\alpha 2$ (A291W), which abolish regulation of the receptor by isoflurane and enflurane (Fig. 4c), did not alter the potentiation of GABA by 1 μ M propofol (wild-type $\alpha 2\beta 1$, $124 \pm 46\%$, $n = 7$; $\alpha 2$ (S270I) $\beta 1$, $88 \pm 29\%$, $n = 7$; $\alpha 2$ (A291W) $\beta 1$, $106 \pm 9\%$, $n = 6$).

Studies of the action of anaesthetics and alcohol on related receptors, such as the serotonin-3 (5-HT₃) and nicotinic acetylcholine receptors (nAChR), have yielded conflicting results concerning their sites of action. Experiments on a single chimaera between the nAChR $\alpha 7$ and 5-HT_{3A} receptor subunits suggest that the N-terminal domain of nAChR $\alpha 7$ is responsible for alcohol¹⁷ and anaesthetic¹⁸ inhibition, but mutations in TM2 increase receptor sensitivity to anaesthetics and alcohols in studies of mouse muscle nAChR subunits¹⁹. In contrast, our data clearly implicate TM2 and TM3. In particular, our result with chimaera C6 shows that this region is not only necessary but also sufficient for the enhancement of receptor function by enflurane and ethanol. Studies using the cysteine substitution mutagenesis method have identified residues in TM2 of the GABA_A-receptor $\alpha 1$ subunit that are accessible from within the aqueous pore. Ser 270 is not among these, and is believed to face directly away from the channel lumen into the hydrophobic interior of the protein²⁰. One possibility suggested, although not proven, by our findings is that Ser 270 and Ala 291 in GABA_A-receptor $\alpha 1$ are part of a binding pocket formed between at least two transmembrane domains. Support for such an idea is provided by the observation that the GABA_A receptor mutations specifically abolished modulation by ethanol and enflurane but did not alter modulation by propofol. Considerable data support the existence within various proteins of sequestered hydrophobic binding sites for small anaesthetic molecules. The existence of such small binding pockets is suggested by the observation of 'cutoff' phenomena in anaesthetic activity as molecular volume is increased^{10,21}. Another possibility is that the anaesthetics bind elsewhere on the glycine and GABA_A receptors and that the residues identified here play some other crucial role in transducing their effects. Physical methods will ultimately be required to resolve this issue beyond reasonable doubt, but high-resolution structures are not yet available for the receptors studied here. Techniques such as X-ray crystallography have already been applied to other proteins of known structure, demonstrating the existence of specific anaesthetic-binding sites between adjacent α -helices in both myoglobin^{22,23} and adenylate kinase²⁴. The creation of receptor mutants such as Gly-receptor $\alpha 1$ (S267I) and GABA_A-receptor $\alpha 2$ (A291W), which lack alcohol or enflurane sensitivity but otherwise function normally, should help

further investigations into the mechanisms of action of volatile anaesthetics and alcohol. The availability of mutants such as these, coupled with transgenic animal technology, now allows the examination of the involvement of specific receptor subunits in the behavioural actions of these compounds. If a mouse expressing a mutated ethanol- or anaesthetic-insensitive receptor were to display altered behavioural sensitivity to those drugs when compared with a wild-type animal, this would provide strong supporting evidence for the role of that receptor in the actions of alcohols or volatile anaesthetics *in vivo*. It would thereby help in distinguishing between nonspecific and specific hypotheses of anaesthetic action. □

Methods

Construction of chimaeras and mutants. The C1 (N-Gly-ALLE/YAAV- ρ -C in TM3), C2 (N-p-SVLE/YAAV-Gly-C in TM3), C4 (N-Gly-AAPA/RVPL- ρ -C in TM1/TM2 loop) and C5 (N-p-AVPA/RVGL-Gly-C in TM1/TM2 loop) chimaeras were constructed by the introduction of restriction sites into homologous regions of the cDNAs encoding the human Gly-receptor $\alpha 1$ and GABA-receptor $\rho 1$ subunits. *Xba*I sites were engineered into the nucleotide sequences of Gly-receptor $\alpha 1$ and GABA-receptor $\rho 1$ corresponding to the region close to the end of TM3 (Fig. 1), and *Bss*HII sites were incorporated into the cDNA sequences encoding the loops between TM1 and TM2 (Fig. 1). Chimaeras were then constructed by standard techniques²⁵. The C3 chimaera (N-p-TLMV/ILSW-Gly-C in TM1) was constructed using the random chimaeragenesis technique²⁶. Gly-receptor $\alpha 1$ and GABA-receptor $\rho 1$ cDNAs were subcloned into the pBK-CMV vector (Stratagene) that had been modified by the removal of the *lac* promoter and the *lacZ* ATG. The GABA-receptor $\rho 1$ and Gly-receptor $\alpha 1$ cDNAs were inserted into this vector in tandem in the *Spe*I and *Clal/Kpn*I sites, respectively. The resulting tandem-containing plasmid was linearized with *Sca*I and *Xba*I; unique restriction sites for these two enzymes were both located between the two parental cDNAs. The linearized DNA was gel purified (QIAEX Resin, Qiagen, Chatsworth, CA) and 200 ng of DNA was used to transform *Escherichia coli* DH5 α (ME; Gibco/BRL). Resultant colonies were screened for the occurrence of homologous recombination events and the presence of GABA-receptor $\rho 1$ /Gly-receptor $\alpha 1$ chimaeras. Finally, chimaera C6 was produced by introducing *Bss*HII sites into the cDNA sequences of chimaeras C1 and C2, and then exchanging fragments in the usual manner. All chimaeras were sequenced (Sequenase version 2.0, U.S. Biochemical) to identify the precise location of the chimaera junctions. Point mutations in GABA-receptor $\alpha 1$, $\alpha 2$, $\beta 1$, $\beta 3$, $\rho 1$ and Gly-receptor $\alpha 1$ subunits were performed on cDNAs subcloned into pCIS2 or pBK-CMV vectors (Stratagene) using either the QuikChange site-directed mutagenesis kit (Stratagene) or the USE kit (Pharmacia Biotech, Piscataway, NJ). All point mutations were verified by double-stranded DNA sequencing.

Electrophysiological recording. *Xenopus* oocytes were isolated and injected²⁷ with human cDNAs (1.5 ng per 30 nl) subcloned into pCIS2 or a modified pBK-CMV vector. Electrophysiological recording was performed as described²⁷. All drugs were applied for 30–180 s and the maximum peaks obtained were measured. To determine the concentration of GABA or glycine producing 10% of the maximal effect (EC₁₀), 1 mM GABA or 10 mM glycine was used to produce a maximal current against which all other concentrations of ligand were compared. The bath concentration of enflurane was determined by gas chromatography^{5,6}. HEK 293 cells (ATCC) were grown on glass coverslips and transfected with cDNAs⁹. Recordings were made in whole-cell mode at –60 mV using patch pipette solutions based on 145 mM *N*-methyl-D-glucamine HCl and extracellular saline based on 145 mM NaCl as described⁹. All drugs were applied locally to the cell using a motor-driven solution exchange device²⁵. Unless otherwise stated, all data in Figs 1, 3 and 4 were obtained in *Xenopus* oocytes using a concentration of agonist (EC₁₀) appropriate for the receptor tested, and are expressed as mean $\pm$ s.e.m. of between 4 and 14 oocytes.

Received 27 February; accepted 1 August 1997.

- Janoff, A. S., Pringle, M. J. & Miller, K. W. Correlation of general anaesthetic potency with solubility in membranes. *Biochim. Biophys. Acta* **649**, 125–128 (1981).
- Lyon, R. C., McComb, J. A., Schreurs, J. & Goldstein, D. B. A relationship between alcohol intoxication and the disordering of brain membranes by a series of short-chain alcohols. *J. Pharmacol. Exp. Ther.* **218**, 669–675 (1981).
- Franks, N. P. & Lieb, W. R. Molecular and cellular mechanisms of general anaesthesia. *Nature* **367**, 607–614 (1994).

4. Hales, T. G. & Lambert, J. J. The actions of propofol on inhibitory amino acid receptors of bovine adrenomedullary chromaffin cells and rodent central neurones. *Br. J. Pharmacol.* **104**, 619–628 (1991).
5. Lin, L.-H., Chen, L. L., Zirrollo, J. A. & Harris, R. A. General anesthetics potentiate gamma-aminobutyric acid actions on gamma-aminobutyric acid receptors expressed in *Xenopus* oocytes: lack of involvement of intracellular calcium. *J. Pharmacol. Exp. Ther.* **263**, 569–578 (1992).
6. Jones, M. V. & Harrison, N. L. Effects of volatile anesthetics on the kinetics of inhibitory postsynaptic currents in cultured rat hippocampal neurons. *J. Neurophysiol.* **70**, 1339–1349 (1993).
7. Mihic, S. J., McQuilkin, S. J., Eger, E. I. II, Ionescu, P. & Harris, R. A. Potentiation of gamma-aminobutyric acid type A receptor-mediated chloride currents by novel halogenated compounds correlates with their abilities to induce general anesthesia. *Mol. Pharmacol.* **46**, 851–857 (1994).
8. Downie, D. L., Hall, A. C., Lieb, W. R. & Franks, N. P. Effects of inhalational general anaesthetics on native glycine receptors in rat medullary neurones and recombinant receptors in *Xenopus* oocytes. *Br. J. Pharmacol.* **118**, 493–502 (1996).
9. Harrison, N. L., Kugler, J. L., Jones, M. V., Greenblatt, E. P. & Pritchett, D. B. Positive modulation of human gamma-aminobutyric acid type A and glycine receptors by the inhalation anesthetic isoflurane. *Mol. Pharmacol.* **44**, 628–632 (1993).
10. Dildy-Mayfield, J. E., Mihic, S. J., Lui, Y., Deitrich, R. A. & Harris, R. A. Actions of long chain alcohols on GABA_A and glutamate receptors: relation to *in vivo* effects. *Br. J. Pharmacol.* **118**, 378–384 (1996).
11. Mascia, M. P., Mihic, S. J., Valenzuela, C. F., Schofield, P. R. & Harris, R. A. A single amino acid determines differences in ethanol actions on strychnine-sensitive glycine receptors. *Mol. Pharmacol.* **50**, 402–406 (1996).
12. Mascia, M. P., Machu, T. K. & Harris, R. A. Enhancement of homomeric glycine receptor function by long-chain alcohols and anaesthetics. *Br. J. Pharmacol.* **119**, 1331–1336 (1996).
13. Ortells, M. O. & Lunt, G. G. Evolutionary history of the ligand-gated ion-channel superfamily of receptors. *Trends Neurosci.* **18**, 121–127 (1995).
14. Mihic, S. J. & Harris, R. A. Inhibition of rho1 receptor GABAergic currents by alcohols and volatile anesthetics. *J. Pharmacol. Exp. Ther.* **277**, 411–416 (1996).
15. Sanna, E., Garau, F. & Harris, R. A. Novel properties of homomeric beta 1 gamma-aminobutyric acid type A receptors: actions of the anesthetics propofol and pentobarbital. *Mol. Pharmacol.* **47**, 213–217 (1995).
16. Jones, M. V., Harrison, N. L., Pritchett, D. B. & Hales, T. G. Modulation of the GABA_A receptor by propofol is independent of the gamma subunit. *J. Pharmacol. Exp. Ther.* **274**, 962–968 (1995).
17. Yu, D. *et al.* Ethanol inhibition of the nicotinic acetylcholine type $\alpha 7$ receptors involves the amino terminal domain of the receptor. *Mol. Pharmacol.* **50**, 1010–1016 (1996).
18. Zhang, L., Oz, M., Stewart, R. R., Peoples, R. W. & Weight, F. F. Volatile general anaesthetic actions on recombinant nACh₁₀, 5-HT₃ and chimeric nACh₁₀-5-HT₃ receptors expressed in *Xenopus* oocytes. *Br. J. Pharmacol.* **120**, 353–355 (1997).
19. Forman, S. A., Miller, K. W. & Yellen, G. A discrete site for general anesthetics on a postsynaptic receptor. *Mol. Pharmacol.* **48**, 574–581 (1995).
20. Xu, M. & Akabas, M. H. Identification of channel-lining residues in the M2 membrane-spanning segment of the GABA-A receptor $\alpha 1$ subunit. *J. Gen. Physiol.* **107**, 195–205 (1996).
21. Franks, N. P. & Lieb, W. R. Mapping of general anaesthetic target sites provides a molecular basis for cutoff effects. *Nature* **316**, 349–351 (1985).
22. Shoenborn, B. P. Binding of cyclopropane to sperm whale myoglobin. *Nature* **214**, 1120–1122 (1967).
23. Shoenborn, B. P., Watson, H. C. & Kendrew, J. C. Binding of xenon to sperm whale myoglobin. *Nature* **207**, 28–30 (1965).
24. Sachsenheimer, W., Pai, E. F., Schulz, G. E. & Schirmer, R. H. Halothane binds in the adenine-specific niche of crystalline adenylate kinase. *FEBS Lett.* **79**, 310–312 (1977).
25. Koltchine, V. V., Ye, Q., Finn, S. E. & Harrison, N. L. Chimeric GABA_A/glycine receptors: expression and barbiturate pharmacology. *Neuropharmacology* **35**, 1445–1456 (1996).
26. Moore, K. R. & Blakely, R. D. Restriction site-independent formation of chimeras from homologous neurotransmitter-transporter cDNAs. *Biotechniques* **17**, 130–135 (1994).
27. Mihic, S. J., Whiting, P. J. & Harris, R. A. Anaesthetic concentrations of alcohols potentiate GABA_A receptor-mediated currents: lack of subunit specificity. *Eur. J. Pharmacol.* **268**, 209–214 (1994).

Acknowledgements. We thank P. Schofield, G. Cutting, P. Whiting and the late D. Pritchett for cDNA; J. L. Kugler and W.-J. Tang for discussions on mutagenesis techniques; T. Dunwiddie, J. Weiner, C. Rick, H. Fozzard and D. McGehee for carefully reading our manuscript; and E. Eger and the UCSF Anesthesia Research Foundation for support of our collaborative project. This research was supported by NIH grants to N.L.H. and R.A.H., a young investigator award from the Foundation for Anesthesia Education and Research (E.P.G.), and an institutional postdoctoral training grant at the University of Chicago (V.V.K.).

Correspondence and requests for materials should be addressed to N.L.H. (e-mail: nh17@midway.uchicago.edu).

Single cyclic nucleotide-gated channels locked in different ligand-bound states

MariaLuisa Ruiz & Jeffrey W. Karpén

Department of Physiology, Department of Physiology and Biophysics, C-240, University of Colorado School of Medicine, Denver, Colorado 80262, USA

Cyclic nucleotide-gated (CNG) channels are directly activated by the binding of several ligands^{1–6}. For these channels as well as for other allosteric proteins, the functional effects of each ligand-binding event have been difficult to assess because ligands continuously bind and unbind at each site. Furthermore, in retinal rod photoreceptors the low cytoplasmic concentration of cyclic GMP⁷ means that channels exist primarily in partially liganded states, so it is important to determine how such channels behave.

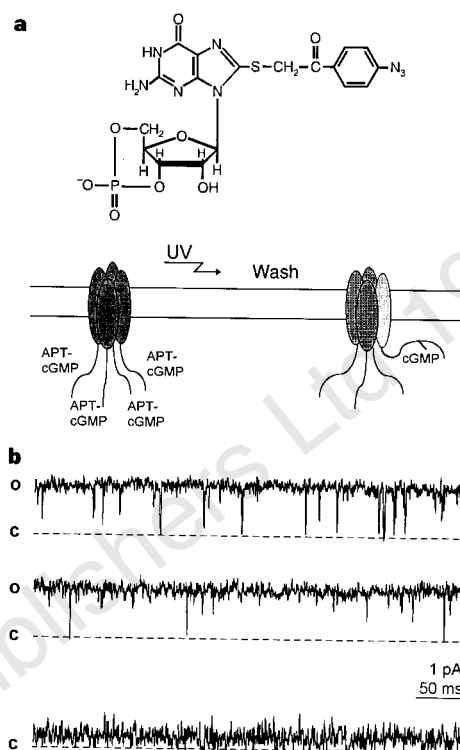


Figure 1 Covalent attachment of ligands persistently activates channels. **a**, Top, structure of APT-cGMP; bottom, representation of APT-cGMP binding and, upon photolysis, covalently attaching cGMP moieties to channel subunits. **b**, Traces from a single channel recording at +50 mV before and after covalent attachment of ligand. Dashed lines indicate closed (C) levels; open, O. Top, maximal activation with 1 mM cGMP; middle, after application of APT-cGMP and exposure to UV light, the patch was washed for 20 min to remove all untethered ligand. Maximal activation persisted in the absence of free cGMP. Bottom, 1 mM Mg²⁺ blocked the persistent current.

Previous studies of single channels have suggested that they occasionally open to subconducting states at low cGMP^{2,3,8–10}, but the significance of these states and how they arise is poorly understood. Here we combine the high resolution of single-channel recording with the use of a photoaffinity analogue of cGMP^{11,12} that tethers cGMP moieties covalently to their binding sites to show single retinal CNG channels can be effectively locked in four distinct ligand-bound states. Our results indicate that channels open more than they would spontaneously when two ligands are bound (~1% of the maximum current), significantly more with three ligands bound (~33%), and open maximally with four ligands bound. In each ligand-bound state, channels opened to two or three different conductance states. These findings place strong constraints on the activation mechanism of CNG channels.

Channels formed from the α -subunit of the retinal rod CNG channel¹³ were expressed in *Xenopus laevis* oocytes and studied in excised membrane patches. This produced homomultimeric channels with four cGMP-binding sites^{14,15}. Patches containing a single channel (see Methods) were exposed to the photoaffinity analogue 8-*p*-azidophenacylthio-cGMP (APT-cGMP¹¹; Fig. 1a) and illuminated with long-wavelength ultraviolet light. Previous work on multichannel patches showed that this procedure resulted in persistent currents that were not removed by extensive perfusion with cGMP-free solution. The persistent currents had several macroscopic properties that precisely matched the behaviour of currents activated by cGMP, indicating that cGMP moieties were covalently tethered to the binding sites^{11,12}. Our goal was to lock single channels into each possible ligand-bound state (Fig. 1a, bottom). Figure 1b compares the behaviour of a single channel fully activated by either cGMP, or by covalently tethered cGMP moieties: in saturating

cGMP, the channel was open most of the time, with an amplitude of about 1.4 pA (top trace); long exposure to ultraviolet light in the presence of APT-cGMP (more than 1.5 min) caused channels to be fully active in the absence of cGMP (middle trace). The current amplitude was unaltered after exposure and the distribution of open times was nearly identical to that observed before covalent attachment of ligand. We next measured block by Mg^{2+} as a test of channel pore integrity: 1 mM Mg^{2+} , applied from the cytoplasmic side, specifically blocks the channel at positive potentials^{11,16–18}. When 1 mM Mg^{2+} was applied to the permanently activated channel, the current was blocked at +50 mV (Fig. 1b, lower trace) but not at –50 mV (not shown). Moreover, the block was flickery¹⁸, in a way identical to that observed in maximally activated channels before exposure to ultraviolet light.

Shorter exposures to ultraviolet light resulted in partially ligand-bound channels. The activation achieved was assessed by measuring dose–response relations before and after exposure. If ligand attached covalently, maximal activation should require lower concentrations of free cGMP. Moreover, the slope of the dose–response relation should be shallower (lower Hill coefficient), as fewer unoccupied sites are available to bind free cGMP. The Hill equation was fitted to the data, and concurrent shifts to lower values in both the $K_{1/2}$ (the cGMP concentration that produces half-maximal activation) and the Hill coefficient were taken as evidence of covalent attachment of ligands. Figure 2 shows superimposed dose–response relations from eight different single-channel patches, measured before and after exposure to ultraviolet light (ten total exposures, where some patches received more than one exposure). The data clustered into four groups, representing control and three distinct shifts in activation. (Fully liganded channels are not considered here, as addition of free ligand had no effect.) These shifts are expressed quantitatively by fitting the Hill equation to the pooled data for each group. Hill coefficients start at 2.9 for the control condition, in which no ligands are covalently attached, and shift to 1.9, 1.4 and 1.1 with progressive exposure to ultraviolet light. As the Hill equation assumes perfect cooperativity of activation, non-integral values are interpreted as lower estimates of the number of simultaneous binding events required for significant activation.

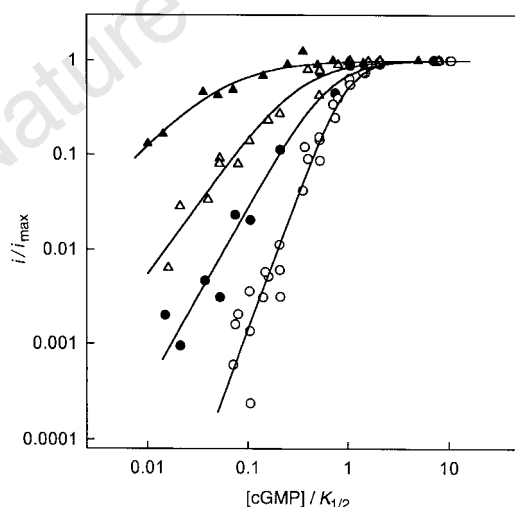


Figure 2 Superimposed cGMP dose–response relations from eight patches, before and after different degrees of covalent ligand attachment (ten total exposures). i/i_{max} , Fractional single-channel current activated by free cGMP. cGMP concentrations are expressed relative to each channel's $K_{1/2}$ (83–160 μ M). This aligns the control dose–response curves, allowing subsequent shifts to be compared between patches¹². The Hill equation (see Methods) fit parameters n and $K_{1/2-p}/K_{1/2}$ are: control (○), 2.9 and 1.0; first shift (●), 1.9 and 0.66 (2 exposures); second shift (△), 1.4 and 0.36 (4 exposures); third shift (▲), 1.1 and 0.055 (4 exposures).

The discrete nature of the dose–response shifts suggests that they are due to the covalent attachment of one, two and three ligands to channel binding sites. These tentative assignments are substantiated by the changes in opening behaviour described below.

We next investigated the channel's tendency to enter into subconducting states. An example of the three conductance states routinely observed (both before and after exposure to ultraviolet light) is shown in Fig. 3. On twenty patches, the mean conductances were: level 1, 10 pS; level 2, 18 pS; level 3, 29 pS. The means varied $\pm 10\%$ between patches. Similar behaviour was reported for native bovine retinal channels⁹, for which it was proposed that each ligand promotes a partial opening of the channel. Others have suggested that overfiltering the records might cut off very fast openings and closings, resulting in a false detection of subconductance states¹⁹. Alternatively, in the related catfish olfactory CNG channel, subconductances were produced by protons binding in the pore^{20,21}. We therefore investigated whether subconductances were linked to channel activation, or could be due to overfiltering or proton block. The three conductances observed in records filtered at 1 kHz were also easily discernible at 5 kHz bandwidth (Fig. 3a). Furthermore, amplitude histograms of the records typically fell into three discrete conductance peaks (Fig. 3b). These observations are inconsistent with inaccurate detection of current levels resulting from overfiltering. In addition, channels tested at pH 8.6 on both sides of the membrane had the same peaks and in proportions similar to those seen at pH 7.6 (data not shown). This suggests that proton block is not responsible for these subconductance states. In fact, block of the native retinal channel by protons has been observed^{22,23}, but with an apparent pK_a of 5.1.

Previous work on multichannel patches suggested that retinal cGMP-gated channels can open spontaneously with very low probability in the absence of ligand^{24,25}. The open probability has recently been estimated²⁵ to be 1.25×10^{-4} for a chimaeric channel comprising the retinal sequence and an olfactory CNG channel pore sequence. To assess the increase in channel open probability caused by ligand binding, we first measured the spontaneous open probability in single channel patches. Spontaneous openings were very brief, and mostly reached only the first and occasionally the second subconductance levels. Three channels were examined for 1.5 to 5 min, and the spontaneous opening probability (open time divided by the total time) averaged 1.5×10^{-5} (s.d. = 8.5×10^{-6}). The degree of spontaneous activity can also be expressed as $i/i_{max} = 6.8 \times 10^{-6}$ (the mean current divided by the maximum current at saturating cGMP).

Figure 4 shows examples of opening behaviour of channels that were singly, doubly, triply and fully liganded in the absence of free cGMP. The assignments are based on the shifts in the dose–response relations in Fig. 2, and are supported by the obvious changes in opening behaviour. Singly liganded channels opened with a very low probability ($i/i_{max} = 9.6 \times 10^{-6}$) that was not significantly different from spontaneous channel opening. This suggests that the binding of one ligand is not sufficient to increase the channel's opening probability. In doubly liganded channels, the probability of opening increased greatly ($i/i_{max} = 0.0097$). Surprisingly, the channel moved freely between two subconductance states and, with low probability, even entered the highest conductance state (Fig. 4, inset). Triply liganded channels opened significantly ($i/i_{max} = 0.33$). They also exhibited the highest probability of opening to subconductance states. A common feature of all partially liganded channels was their preference for opening to subconductance states over the highest conductance state. In contrast, fully liganded channels favoured opening to the highest conductance state; the probability of opening to lower conductance states fell below that in triply liganded channels.

The dependence of subconducting states on the level of activation is not a consequence of covalent attachment of ligands. The

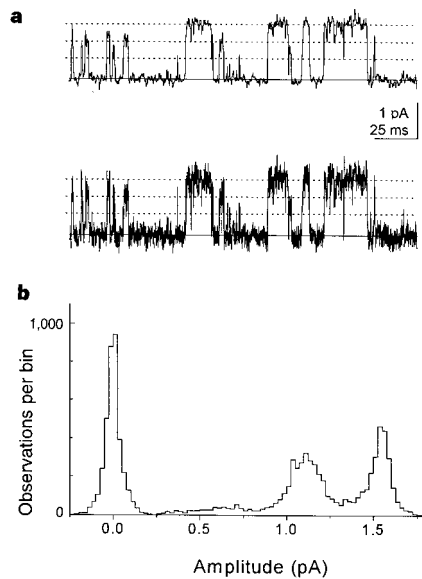


Figure 3 Subconducting cGMP concentrations produce subconducting states. **a**, Single channel at +50 mV, activated with 50 μ M cGMP. The same current record was filtered at 1 kHz and sampled at 5 kHz (top), or filtered at 5 kHz and sampled at 25 kHz (bottom). Solid line, closed level; three dashed lines, conductance levels. **b**, Amplitude histogram of the same record at 1 kHz (total time was 1.5 min).

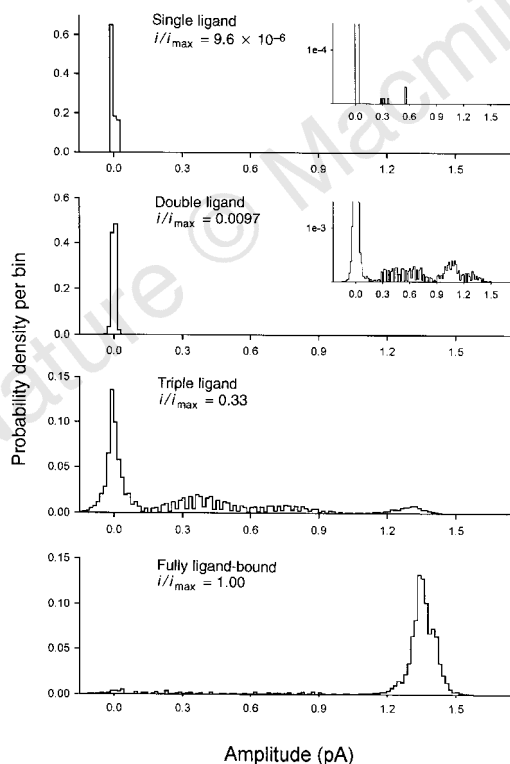


Figure 4 Opening behaviour of partially and fully ligand-bound single channels. The number of ligands covalently attached is labelled at the top of each frame. Probability density histograms illustrate the proportion of time spent in each current level. Insets show the data on an expanded scale to allow examination of conductance state occupancy when the probability of opening was very low. The persistent activation of channels is represented by $i/i_{\max}$ (see text). For eleven data sets, channels occupied by the same number of ligands varied in $i/i_{\max}$ up to 5-fold (much less than the differences between liganded states), and the abundance of any given subconductance state also varied by a similar factor.

channel's probability of opening to either subconductance state peaked at subsaturating concentrations of free cGMP (near $K_{1/2}$). (See also refs 8–10). Furthermore, after covalent attachment of two ligands, the occupancy of both subconductance states increased further with the addition of cGMP and decreased with saturating cGMP (data not shown). Overall, these findings demonstrate that subconductance states are a fundamental part of the channel's mechanism of activation.

In conclusion, we have locked channels into partially liganded states and studied their behaviour without the ambiguities that arise from continuous binding and unbinding of ligands to channel subunits. This level of resolution allowed us to observe all of the steps leading to complete activation. Surprisingly, partially liganded channels exhibit a dynamic flexibility in opening that contrasts with the simple closed–open transition postulated in concerted allosteric models. It has been proposed^{6,15,25} that the Monod–Wyman–Changoux model²⁶ could describe the activation of this channel. Two of our findings indicate that any such concerted model is inadequate: (1) Multiple open states are associated with each liganded state; and (2) for any given open state, the channel-opening equilibrium constant does not increase by a constant factor for each ligand that binds. The first finding also argues against a model in which each binding event is associated with a distinct conductance state⁹. However, the probability of opening to any given conducting state does depend on the number of ligands bound. How the subunits work together to give rise to different conducting states remains to be determined. The information gained from our approach, in conjunction with structural studies, will allow formulation of a more precise model of channel activation. □

Methods

Xenopus laevis oocytes were injected with cRNA (0.046–0.23 ng) transcribed from cloned bovine retina cDNA encoding the α -subunit of the rod CNG channel. Incubation was at 16 °C for 2–5 days. Inside-out membrane patches were excised with patch electrodes (8–23 M Ω) coated with Sylgard and filled with control solution containing 500 μ M niflumic acid. Control solution contained (mM): 130 NaCl, 2 HEPES, 0.02 EDTA, 1 EGTA, pH 7.6 (or 8.6) with NaOH. cGMP was from Sigma. Patches containing a single channel, as judged by a maximum current of about 1.4 pA at +50 mV in saturating cGMP, were analysed further. APT-cGMP was synthesized in our laboratory¹¹. For covalent activation, patches bathed in 20 μ M APT-cGMP (a nearly saturating concentration) were illuminated with light centred at 360 nm from a 200-W mercury lamp¹². Exposures were from 10 s to 3 min. After exposure, patches were washed for 10–30 min in control solution to remove all untethered ligand¹² before dose–response assays were performed. Exposure to UV light alone did not alter channel behaviour: in three patches, dose–response relations were identical before and after exposure. Channel activity at each cGMP concentration was recorded for a minimum of 30 s. Currents at +50 mV were acquired with an Axopatch 200A amplifier (Axon Instruments), filtered at 5 kHz with an 8-pole Bessel filter, digitized at 88 kHz and stored on tape with a Neuro-corder DR-484 PCM recording unit (Neuro Data Instruments). For analysis, most of the data was resampled at 5 kHz and filtered at 1 kHz. Probability density histograms and dose–response relations were plotted in Sigma Plot (Jandel Scientific) from events lists constructed with Fetchan and Pstat (Pclamp 6, Axon Instruments). A modified Hill equation¹² was fitted to dose–response data: $i/i_{\max} = ([\text{cGMP}]/K_{1/2})^n / \{([\text{cGMP}]/K_{1/2})^n + (K_{1/2-p}/K_{1/2})^n\}$, where n is the Hill coefficient and $K_{1/2-p}/K_{1/2}$ is the ratio of the $K_{1/2}$ after covalent ligand attachment to the original $K_{1/2}$. Long closed times of the order of hundreds of milliseconds to several seconds^{9,10,18,27} occurred at low frequency and were removed from analysis on the basis of stability plots.

Received 13 May; accepted 10 July 1997.

1. Fesenko, E. E., Kolesnikov, S. S. & Lyubarsky, A. L. Induction by cyclic GMP of cationic conductance in plasma membrane of retinal rod outer segment. *Nature* **313**, 310–313 (1985).
2. Haynes, L. W., Kay, A. R. & Yau, K.-W. Single cyclic GMP-activated channel activity in excised patches of rod outer segment membrane. *Nature* **321**, 66–70 (1986).
3. Zimmerman, A. L. & Baylor, D. A. Cyclic GMP-sensitive conductance of retinal rods consists of aqueous pores. *Nature* **321**, 70–72 (1986).
4. Nakamura, T. & Gold, G. H. A cyclic nucleotide-gated conductance in olfactory receptor cilia. *Nature*

- 325, 442–444 (1987).
5. Zimmerman, A. L. Cyclic nucleotide-gated channels. *Curr. Opin. Neurobiol.* **5**, 296–303 (1995).
6. Zagotta, W. N. & Siegelbaum, S. A. Structure and function of cyclic nucleotide-gated channels. *Annu. Rev. Neurosci.* **19**, 235–263 (1996).
7. Nakatani, K. & Yau, K.-W. Guanosine 3',5'-cyclic monophosphate-activated conductance studied in a truncated rod outer segment of the toad. *J. Physiol.* **395**, 731–753 (1988).
8. Hanke, W., Cook, N. J. & Kaupp, U. B. cGMP-dependent channel protein from photoreceptor membranes: single channel activity of the purified and reconstituted protein. *Proc. Natl Acad. Sci. USA* **85**, 94–98 (1988).
9. Ildefonso, M. & Bennett, N. Single channel study of the cGMP-dependent conductance of retinal rods from incorporation of native vesicles into planar lipid bilayers. *J. Membr. Biol.* **123**, 133–147 (1991).
10. Taylor, W. R. & Baylor, D. A. Conductance and kinetics of single cGMP-activated channels in salamander rod outer segments. *J. Physiol.* **483**, 567–582 (1995).
11. Brown, R. L., Gerber, W. V. & Karpen, J. W. Specific labeling and permanent activation of the retinal rod cGMP-activated channel by the photoaffinity analog 8-*p*-azidophenacylthio-cGMP. *Proc. Natl Acad. Sci. USA* **90**, 5369–5373 (1993).
12. Karpen, J. W. & Brown, R. L. Covalent activation of retinal rod cGMP-gated channels reveals a functional heterogeneity in the ligand binding sites. *J. Gen. Physiol.* **107**, 169–181 (1996).
13. Kaupp, U. B. *et al.* Primary structure and functional expression from complementary DNA of the rod photoreceptor cyclic GMP-gated channel. *Nature* **342**, 762–766 (1989).
14. Liu, D. T., Tibbs, G. R. & Siegelbaum, S. A. Subunit stoichiometry of cyclic nucleotide-gated channels and effects of subunit order on channel function. *Neuron* **16**, 983–990 (1996).
15. Varnum, M. D. & Zagotta, W. N. Subunit interactions in the activation of cyclic nucleotide-gated ion channels. *Biophys. J.* **70**, 2667–2679 (1996).
16. Colamartino, G., Menini, A. & Torre, V. Blockage and permeation of divalent cations through the cyclic GMP-activated channel from tiger salamander retinal rods. *J. Physiol.* **440**, 189–206 (1991).
17. Zimmerman, A. L. & Baylor, D. A. Cation interactions within the cyclic GMP-activated channel of retinal rods from the tiger salamander. *J. Physiol.* **449**, 759–783 (1992).
18. Nizzari, M. *et al.* Single-channel properties of cloned cGMP-activated channels from retinal rods. *Proc. R. Soc. Lond. B* **254**, 69–74 (1993).
19. Torre, V., Straforini, M., Sesti, F. & Lamb, T. D. Different channel-gating properties of two classes of cyclic GMP-activated channel in vertebrate photoreceptors. *Proc. R. Soc. Lond. B* **250**, 209–215 (1992).
20. Goulding, E. H. *et al.* Molecular cloning and single-channel properties of the cyclic nucleotide-gated channel from catfish olfactory neurons. *Neuron* **8**, 45–58 (1992).
21. Root, M. J. & MacKinnon, R. Two identical noninteracting sites in an ion channel revealed by proton transfer. *Science* **265**, 1852–1856 (1994).
22. Tanaka, J. C. The effects of protons on 3',5'-cGMP-activated currents in photoreceptor patches. *Biophys. J.* **65**, 2517–2523 (1993).
23. Picco, C., Sanfilippo, C., Gavazzo, P. & Menini, A. Modulation by internal protons of native cyclic nucleotide-gated channels from retinal rods. *J. Gen. Physiol.* **108**, 265–276 (1996).
24. Picones, A. & Korenbrot, J. I. Spontaneous, ligand-independent activity of the cGMP-gated ion channels in cone photoreceptors of fish. *J. Physiol.* **485**, 699–714 (1995).
25. Tibbs, G. R., Goulding, E. H. & Siegelbaum, S. A. Allosteric-activation and tuning of ligand efficacy in cyclic-nucleotide-gated channels. *Nature* **386**, 612–615 (1997).
26. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
27. Matthews, G. & Watanabe, S.-I. Activation of single ion channels from toad retinal rod inner segments by cyclic GMP: concentration dependence. *J. Physiol.* **403**, 389–405 (1988).

Acknowledgements. We thank W. Betz, R. L. Brown, Y. Koutalos, A. R. Martin and W. Sather for comments on the manuscript. This work was supported by a grant from the National Eye Institute. M.L.R. was the recipient of a National Research Service Award.

Correspondence and requests for materials should be addressed to J.W.K. (e-mail: jeffrey.karpen@uchsc.edu).

G-protein deactivation is rate-limiting for shut-off of the phototransduction cascade

Mandeep S. Sagoo & Leon Lagnado*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Photoreceptors detect light through a seven-helix receptor (rhodopsin) and heterotrimeric G protein (transducin) coupled to a cyclic GMP phosphodiesterase^{1,2}. Similar pathways are used to amplify responses to hormones, taste and smell^{3–5}. The amplification of phototransduction is reduced by a fall in cytoplasmic Ca²⁺ (refs 6–10), but it is not known how the deactivation of rhodopsin and transducin influence this response and hence the extent and duration of phosphodiesterase activity^{11–14}. Here we investigate this by recording the electrical response to flashes of light in truncated rod photoreceptors¹⁰. By removing ATP to block the deactivation of rhodopsin by phosphorylation¹⁵, we show that this reaction limits the amplitude of the response and begins within 3.2 s of a flash in a solution containing 1 μM Ca²⁺, falling to 0.9 s in a zero-Ca²⁺ solution. In contrast, the activation and amplitude of the response were unaffected when transducin deactivation by GTP hydrolysis was blocked by replacing GTP with its non-

hydrolysable analogue GTP-γS¹¹, demonstrating that there is little GTP hydrolysis occurring over the period in which photo-excited rhodopsin is quenched. The rapid deactivation of rhodopsin is therefore a Ca²⁺-sensitive step controlling the amplitude of the light response, whereas transducin deactivation is slower and controls recovery.

The first stage of amplification in phototransduction occurs when a single molecule of photoexcited rhodopsin (Rh*) activates hundreds of copies of transducin^{1–3}, but it is not clear how the reactions involved in the deactivation of rhodopsin and transducin affect this response¹⁶. We therefore recorded the cGMP-activated current in truncated salamander rods that had been internally perfused with solutions that could be rapidly changed^{8,10}. We first tested how the quenching of Rh* affects the electrical response generated when cGMP is hydrolysed by light-activated phosphodiesterase (PDE). Rh* needs to be phosphorylated at several sites by rhodopsin kinase for deactivation¹⁵, which can therefore be blocked by removing ATP from the perfusate. Figure 1a shows a control response to a sub-saturating flash of light delivered in the presence of 1 mM ATP (thin trace), and a test response to a flash of the same strength delivered after removing ATP (bold trace). In both cases, the perfusate contained zero Ca²⁺, as well as 5 μM GTP to allow activation of transducin (T*). The control response reached 60% of maximum and began to recover within 3 s, but the test response reached saturation and there was very little recovery until 1 mM ATP was added. Examination of the rising phases on an expanded timescale (Fig. 1b) demonstrated that removing ATP did not affect the initial rise in PDE activity, but after a delay of 0.55 s the response in 1 mM ATP rose more slowly, indicating that the quenching of Rh* had begun to limit the activation of transducin and PDE. This behaviour was observed in all 25 cells tested and demonstrates that Rh* deactivation severely limits the amplitude of the flash response. This idea is supported by experiments on rods from transgenic mice containing a fraction of rhodopsin molecules in which the phosphorylation sites have been truncated¹⁷.

The rate of decline of PDE activity back to basal levels after a flash will be determined by whichever of the two preceding members of the cascade, Rh* or T*, deactivates more slowly. Both Rh* (refs 12, 18, 19) and T* (ref. 20) have been suggested to be the longer-lived species that controls recovery. Rh* deactivation is rapid enough to limit the amplitude of the flash response (Fig. 1), so if T* deactivation were faster it should also limit the response. T* deactivation occurs when the G-protein α-subunit hydrolyses its bound GTP, a reaction that can be blocked by replacing GTP with GTP-γS¹¹. Figure 2a shows the responses to sub-saturating flashes of fixed strength delivered in 5 μM GTP and 5 μM GTP-γS, both in 1 μM Ca²⁺. The control response recovered fully, but the test response showed only a small recovery before remaining steady for over 10 s, confirming that almost all T* deactivation by GTP hydrolysis was blocked. Figure 2b shows the rising phases of these responses on an expanded timescale. Blocking GTP hydrolysis did not affect the rate of activation nor, crucially, the peak amplitude of the response. T* deactivation by GTP hydrolysis must therefore be considerably slower than Rh* deactivation.

When GTP hydrolysis is blocked, the amount of T* (and hence the amplitude of the response) is expected to grow as long as Rh* remains unquenched and continues to interact with transducin¹¹. The fact that the amplitude of the response was not increased by blocking T* deactivation with GTP-γS therefore indicates that the creation of T* does not continue to any appreciable extent after the peak of the response. Figures 1 and 3 show that this limit to the amplitude of the flash response was not set by the supply of transducin, as might be expected if a single Rh* activated most of the transducin present on the disc membrane, because removing ATP allowed the amplitude of the response to increase. We therefore conclude that the amplitude of the response is limited by deactivation of Rh*, which is effectively complete at the peak of

the flash response, and that little deactivation of T^* occurs over this time. GTP hydrolysis by transducin is therefore the rate-limiting reaction in turning off PDE activity.

The amplification of the phototransduction cascade is reduced by light itself, and the signal for this process of adaptation is the fall in cytoplasmic Ca^{2+} that accompanies the closure of cGMP-gated channels^{6,7,21}. A key question has been the identity of the Ca^{2+} -sensitive step(s) controlling the gain of phototransduction. The flash response reaches its peak sooner in a light-adapted rod, and this observation has been interpreted to indicate that the reaction controlling recovery is accelerated (reviewed in ref. 16). We tested this idea in the experiment shown in Fig. 2c. The thin traces show a flash of fixed strength delivered in $1 \mu M$ Ca^{2+} , when the response saturated, and in zero Ca^{2+} , when it was desensitized. When T^* deactivation was blocked by replacing GTP with GTP- γS , the response in zero Ca^{2+} was desensitized to the same degree (bold trace) and the activation phases superimposed (Fig. 2d), even though recovery was completely blocked. T^* deactivation was therefore rate-limiting for recovery in both $1 \mu M$ Ca^{2+} (Fig. 2a; 4 cells) and zero Ca^{2+} (Fig. 2c; 3 cells), and did not participate in gain control by Ca^{2+} . These results are consistent with previous experiments indicating that the reaction controlling recovery is not

sensitive to Ca^{2+} (refs 13, 14).

Several experiments indicate that the action of Ca^{2+} on the amplification of the phototransduction cascade is exerted only over a narrow period during or shortly after the creation of Rh^* (refs 10, 14, 22). As Rh^* deactivation began rapidly (Fig. 1) and was complete by the time the response reached its peak (Fig. 2), we tested whether this was a Ca^{2+} -sensitive reaction involved in gain control. To detect changes in the rate of Rh^* deactivation, we measured the delay between the delivery of a flash and the first measurable effects of Rh^* deactivation by comparing the rising phases of responses in the presence and absence of ATP as described for Fig. 1. Figure 3a shows the activation of flash responses in $1 \mu M$ Ca^{2+} , first in $1 mM$ ATP (thin trace) and then after Rh^* deactivation was blocked in zero ATP (bold trace). The delay between the flash and the peeling away of the two responses was 1.9 s. In zero Ca^{2+} , the response was desensitized and the delay before the effects of Rh^* deactivation could be detected was shortened to 1 s (Fig. 3b). The decrease in sensitivity and accelerated deactivation of Rh^* could not be reversed by returning to $1 \mu M$ Ca^{2+} , when the effects of Rh^* deactivation could be detected within 0.8 s of a flash (Fig. 3c). The delay between the creation of Rh^* and the beginning of measurable deactivation was 3.2 ± 0.6 s in $1 \mu M$ Ca^{2+} (5 cells), falling to

Figure 1 Quenching of photoexcited rhodopsin is apparent after a short delay. **a**, Normalized responses to a flash of fixed strength delivered in the presence (thin trace) and absence (bold trace) of $1 mM$ ATP. Flash ($15 \text{ photons } \mu m^{-2}$) delivered at 0 s (upward-pointing arrow). The control response reached 60% of maximum, but the test response saturated until after 9 s $1 mM$ ATP was added back to the perfusate (downward arrow) to allow recovery. **b**, Rising phases of the responses in **a** shown on an expanded timescale. The inhibitory effect of Rh^* phosphorylation was apparent after a delay of 0.55 s. Concentrations were $5 \mu M$ GTP, $0 Ca^{2+}$ throughout. Note that $1 mM$ ATP did not increase the initial rate of activation of the response, indicating that the GTP concentration inside the outer segment was constant (see Methods).

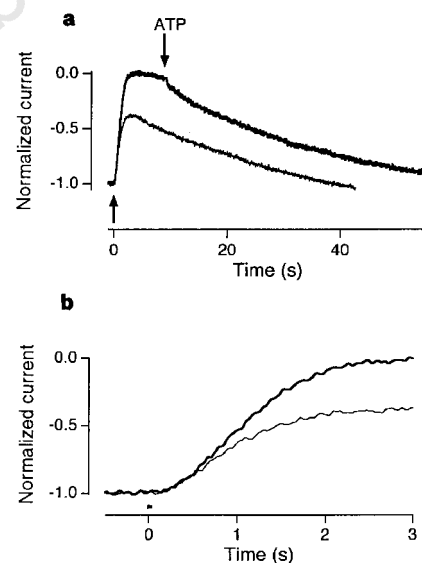
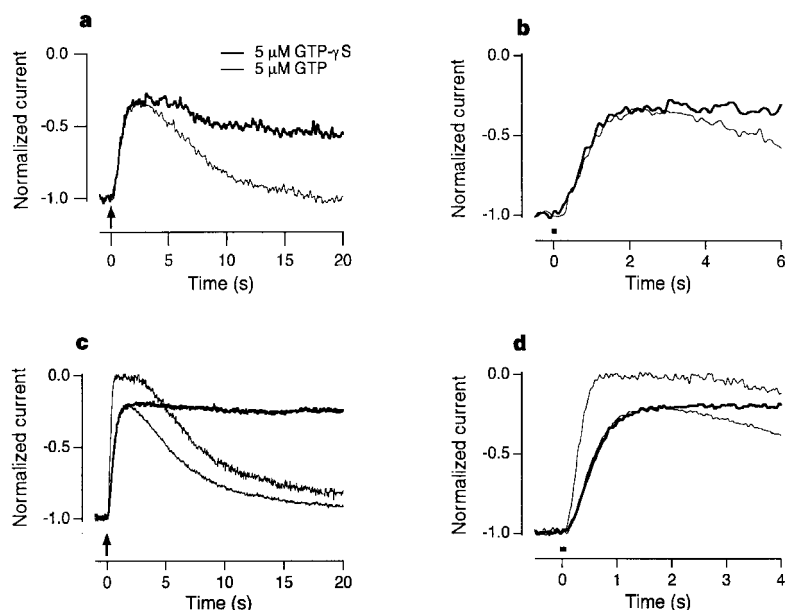


Figure 2 Quenching of photoexcited rhodopsin is effectively complete at the peak of the flash response. **a**, Normalized responses to a flash of fixed strength delivered in the presence of $5 \mu M$ GTP (thin trace) or $5 \mu M$ GTP- γS (bold trace). Flash ($10 \text{ photons } \mu m^{-2}$) delivered at 0 s (lower arrow). Both responses reach 70% of maximum, but the response in GTP- γS shows only a small recovery, confirming that most transducin deactivation by GTP hydrolysis was blocked. **b**, Rising phases of the responses in **a** shown on an expanded timescale. The responses activate at the same rate, indicating that GTP- γS and GTP activate transducin with similar efficiency. Note that the response in GTP- γS is not significantly larger. Concentrations were $1 mM$ ATP and $1 \mu M$ Ca^{2+} throughout. **c**, Thin traces show responses to a flash of fixed strength ($19 \text{ photons } \mu m^{-2}$) delivered in $1 \mu M$ Ca^{2+} and $0 Ca^{2+}$. The response in $0 Ca^{2+}$ has a smaller amplitude, showing the desensitizing effect of lowering Ca^{2+} . The response in $0 Ca^{2+}$ was desensitized equally when $5 \mu M$ GTP- γS replaced GTP (bold trace). **d**, Rising phases of the responses in **c** shown on an expanded timescale. Blocking T^* deactivation did not affect the activation of the response.



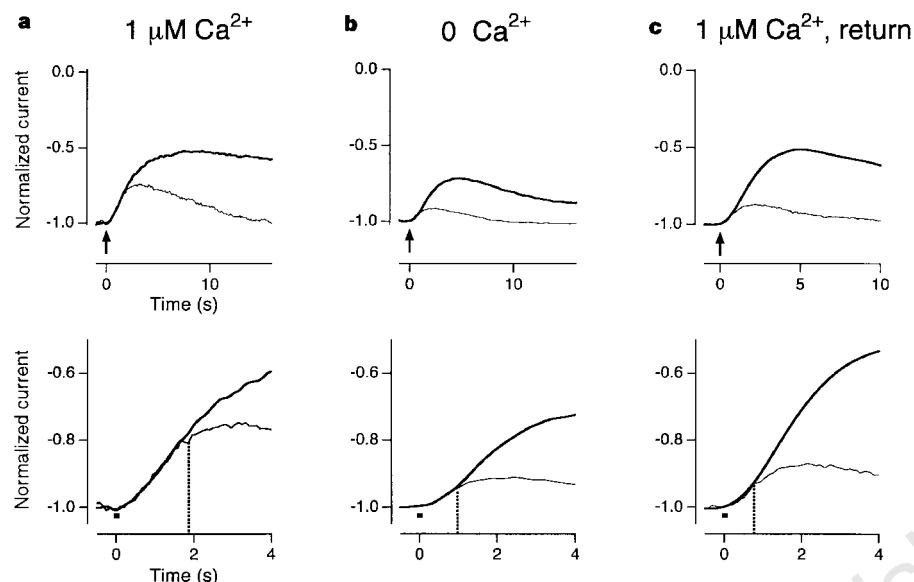


Figure 3 Removing Ca^{2+} accelerates the quenching of photoexcited rhodopsin. Responses are shown to a flash of fixed strength ($9 \text{ photons } \mu\text{m}^{-2}$), delivered in $1 \mu\text{M } \text{Ca}^{2+}$ (**a**), then 0 Ca^{2+} (**b**), then again in $1 \mu\text{M } \text{Ca}^{2+}$ (**c**). In each case the thin trace is a response in 1 mM ATP , and the bold trace is a response in 0 ATP . Lower panels show the rising phases on an expanded timescale and the times at which the two responses diverged were estimated to be 1.9 s (**a**), 1 s (**b**) and 0.8 s (**c**). The first pair of responses in $1 \mu\text{M } \text{Ca}^{2+}$ in **a** were obtained within 2 min of truncation.

$0.9 \pm 0.1 \text{ s}$ in 0 Ca^{2+} (9 cells). In the intact rod outer segment, the free Ca^{2+} concentration in the dark is about $500\text{--}600 \text{ nM}$ (refs 23, 24), falling in the light to an estimated level of between 50 nM (ref. 24) and close to zero^{25,26}. The deactivation of Rh^* was therefore accelerated by a fall in Ca^{2+} spanning the physiological range.

The irreversible decrease in sensitivity after exposure to zero Ca^{2+} (Fig. 3) is consistent with experiments in truncated rods^{8,10}, and has been attributed to the loss of a soluble protein involved in controlling sensitivity⁹. This was identified as the Ca^{2+} -binding protein S-modulin (also known as recoverin²⁷), which inhibited phosphorylation of Rh^* (ref. 9). S-modulin is therefore a candidate for the Ca^{2+} -sensitive molecule controlling the rapid deactivation of Rh^* shown in Fig. 3.

These results indicate that the response generated by the creation of Rh^* following a flash of light consists of two distinct phases that do not overlap significantly: first, transducin molecules are activated by Rh^* , then, after Rh^* is quenched, these active G proteins slowly deactivate by GTP hydrolysis. Ca^{2+} controls the rapid deactivation of Rh^* to regulate the number of T^* created by each Rh^* . This is probably not the only mechanism of gain control, however, because light adaptation also involves a Ca^{2+} -dependent decrease in the initial rate of activation which is functionally equivalent to creating fewer Rh^* (refs 10, 28, 29). Nonetheless, our results support the proposal that distinct steps control the gain of the phototransduction cascade and the time constant of recovery^{13,14,22}. In salamander rods, recovery after a flash occurs with a dominant time constant of 2.4 s (refs 12, 14), which is independent of Ca^{2+} (ref. 13) and can now be attributed to GTP hydrolysis by transducin. When Ca^{2+} in the rod is clamped at the higher levels that occur in darkness, recovery of the flash response can be described by a model in which a second species decays with a time constant of 0.4 s (ref. 13). Our results indicate that this species is Rh^* , and a time constant of 0.4 s is consistent with the observation that Rh^* deactivation is effectively complete within $2\text{--}3 \text{ s}$ of a flash delivered in $1 \mu\text{M } \text{Ca}^{2+}$, and before any appreciable deactivation of transducin (Fig. 2). Our results predict that this second time constant will become shorter when Ca^{2+} inside the rod outer segment falls during the normal response to light.

One advantage of sequential, rather than simultaneous, activation and deactivation of G proteins following activation of a receptor is that a particular level of enzyme activity is achieved with minimum expenditure of GTP. Gain control by regulation of the active lifetime of the receptor, rather than the rate of GTP hydrolysis by the G protein, maintains this high level of efficiency. It

will be interesting to see whether gain control at the receptor and timescale control at the G protein are general features of G-protein-mediated responses. □

Methods

Recording from truncated rods. Suction electrodes were used to record the cGMP-activated current in the outer segments of truncated rods from the retinae of larval tiger salamanders (*Ambystoma tigrinum*) using methods described previously¹⁰. Briefly, a rod was drawn, outer segment first, into a suction pipette which collected a fraction of the current flowing across the outer-segment plasma membrane. The inner segment of the rod was then amputated using a glass probe, leaving the cut end of the outer segment open to the bath. The truncated rod was held by one mouth of a glass theta-tube which allowed the solution perfusing the inside of the outer segment to be switched in about 20 ms . Light flashes lasted 20 ms and were at 500 nm wavelength (8 nm half-height bandwidth). All experiments were done in darkness at room temperature ($18\text{--}20^\circ\text{C}$).

Solutions. Cells were isolated in Ringer solution containing (in mM): NaCl, 110; CaCl_2 , 1; MgCl_2 , 1.6; KCl, 2.5; HEPES, 10; glucose, 5; pH adjusted to 7.6 with NaOH. The suction electrode contained (in mM): NaCl, 112.5; HEPES, 5; MgCl_2 , 2; EGTA, 0.1. The standard solution perfusing the inside of the outer segment contained (in mM): arginine glutamate, 110; MgCl_2 , 2; HEPES, 5; EGTA, 1; CaCl_2 , 0.985; cGMP, 0.1; ATP, 1; GTP, 0.005; pH adjusted to 7.7 with NaOH. The calculated free $[\text{Ca}^{2+}]$ in this solution was 1 mM , assuming the K_d of EGTA under these conditions is 16.75 nM . The actual free $[\text{Ca}^{2+}]$ in solutions prepared in this way was checked using a Ca^{2+} -selective electrode, calibrated with Ca^{2+} standards (Molecular Probes). Solutions containing nominally zero Ca^{2+} were made by omitting CaCl_2 . All chemicals were obtained from Sigma, except EGTA (Fluka) and $1 \text{ M } \text{CaCl}_2$ stock solution (BDH).

The initial rising phase of the flash response in a truncated rod is dependent on the GTP concentration, with a $K_{1/2}$ of $9 \mu\text{M}$ GTP in $1 \mu\text{M } \text{Ca}^{2+}$ (ref. 30). The constancy of the initial rising phase on adding 1 mM ATP (Figs 1 and 3) therefore argues against the possibility that the GTP concentration inside the outer segment was significantly increased by, for instance, the introduction of a GTP impurity or phosphorylation of GMP. The cGMP-activated current recorded in darkness was also relatively constant on addition of 1 mM ATP , confirming that there was no appreciable stimulation of cGMP synthesis by the guanylyl cyclase.

Received 1 May; accepted 7 July 1997.

1. Stryer, L. The cyclic nucleotide cascade of vision. *Annu. Rev. Neurosci.* **9**, 87–119 (1986).
2. Lagnado, L. & Baylor, D. Signal flow in visual transduction. *Neuron* **8**, 995–1002 (1992).
3. Lamb, T. D. & Pugh, E. N. G-protein cascades: gain and kinetics. *Trends Neurosci.* **15**, 291–298 (1992).
4. Kolesnikov, S. S. & Margolskee, R. F. A cyclic-nucleotide-suppressible conductance activated by transducin in taste cells. *Nature* **376**, 85–88 (1995).
5. Gold, G. H. & Pugh, E. N. The nose leads the eye. *Nature* **385**, 677–679 (1997).
6. Matthews, H. R., Murphy, R. L., Fain, G. L. & Lamb, T. D. Photoreceptor light adaptation is mediated

- by cytoplasmic calcium concentration. *Nature* **334**, 67–69 (1988).
7. Nakatani, K. & Yau, K.-W. Calcium and light adaptation in retinal rods and cones. *Nature* **334**, 69–71 (1988).
 8. Kawamura, S. & Murakami, M. Calcium-dependent regulation of cyclic GMP phosphodiesterase by a protein from frog retinal rods. *Nature* **349**, 420–423 (1991).
 9. Kawamura, S. Rhodopsin phosphorylation as a mechanism of cyclic GMP phosphodiesterase regulation by S-modulin. *Nature* **362**, 855–857 (1993).
 10. Lagnado, L. & Baylor, D. A. Calcium controls light-triggered formation of catalytically active rhodopsin. *Nature* **367**, 273–277 (1994).
 11. Vuong, T. M. & Chabre, M. Deactivation kinetics of the transduction cascade of vision. *Proc. Natl Acad. Sci. USA* **88**, 9813–9817 (1991).
 12. Pepperberg, D. R. et al. Light-dependent delay in the falling phase of the retinal rod photoresponse. *Vis. Neurosci.* **8**, 9–18 (1992).
 13. Lyubarsky, A., Nikonov, S. & Pugh, E. N. The kinetics of inactivation of the rod phototransduction cascade with constant Ca^{2+} . *J. Gen. Physiol.* **107**, 19–34 (1996).
 14. Murnick, J. G. & Lamb, T. D. Kinetics of desensitization induced by saturating flashes in toad and salamander rods. *J. Physiol. (Lond.)* **495**, 1–13 (1996).
 15. Wilden, U., Hall, S. W. & Kuhn, H. Phosphodiesterase activation by photoexcited rhodopsin is quenched when rhodopsin is phosphorylated and binds the intrinsic 48 kD protein of rod outer segments. *Proc. Natl Acad. Sci. USA* **83**, 1174–1178 (1986).
 16. Detwiler, P. B. & Gray-Keller, M. P. The mechanisms of vertebrate light adaptation: speeded recovery versus slowed activation. *Curr. Opin. Neurobiol.* **6**, 440–444 (1996).
 17. Chen, J., Makino, C. L., Peachey, N. S., Baylor, D. A. & Simon, M. I. Mechanisms of rhodopsin inactivation *in vivo* as revealed by a COOH-terminal truncation mutant. *Science* **267**, 374–377 (1995).
 18. Pepperberg, D. R., Jin, J. & Jones, G. J. Modulation of transduction gain in light adaptation of retinal rods. *Vis. Neurosci.* **11**, 53–62 (1994).
 19. Corson, D. W., Cornwall, M. C. & Pepperberg, D. R. Evidence for the prolonged photoactivated lifetime of an analogue visual pigment containing 11-*cis* 9-desmethylretinal. *Vis. Neurosci.* **11**, 91–98 (1994).
 20. Nikonov, S. S., Lyubarsky, A. L. & Pugh, E. N. Hydroxylamine decreases rod sensitivity *in situ* without changing the dominant time constant of recovery. *Invest. Ophthalmol. Vis. Sci.* **37**, S5 (1996).
 21. McNaughton, P. A., Cervetto, L. & Nunn, B. J. Measurement of the intracellular free calcium concentration in salamander rods. *Nature* **322**, 261–263 (1986).
 22. Matthews, H. R. Static and dynamic actions of cytoplasmic Ca^{2+} in the adaptation of responses to saturating flashes in salamander rods. *J. Physiol. (Lond.)* **490**, 1–15 (1996).
 23. Lagnado, L., Cervetto, L. & McNaughton, P. A. Calcium homeostasis in the outer segments of retinal rods from the tiger salamander. *J. Physiol. (Lond.)* **455**, 111–142 (1992).
 24. Gray-Keller, M. P. & Detwiler, P. B. The calcium feedback signal in the phototransduction cascade of vertebrate rods. *Neuron* **13**, 849–861 (1994).
 25. Cervetto, L., Lagnado, L., Perry, R. J., Robinson, D. W. & McNaughton, P. A. Extrusion of calcium from rod outer segments is driven by both sodium and potassium gradients. *Nature* **337**, 740–743 (1989).
 26. Younger, J. P., McCarthy, S. T. & Owen, W. G. Modulation of the cytosolic free calcium and changes in sensitivity and circulating current occur over the same range of steady-state adapting lights in rod photoreceptors. *Invest. Ophthalmol. Vis. Sci.* **33**, 1104 (1992).
 27. Tanaka, T., Ames, J. B., Harvey, T. S., Stryer, L. & Ikura, M. Sequestration of the membrane-targeting myristoyl group of recoverin in the calcium-free state. *Nature* **376**, 444–447 (1995).
 28. Jones, G. J. Light adaptation and the rising phase of the flash photocurrent of salamander retinal rods. *J. Physiol. (Lond.)* **487**, 441–451 (1995).
 29. Gray-Keller, M. P. & Detwiler, P. B. Ca^{2+} dependence of dark- and light-adapted flash responses in rod photoreceptors. *Neuron* **17**, 323–331 (1996).
 30. Sagoo, M. S. & Lagnado, L. Ca^{2+} increases the apparent affinity of transducin for GTP. *Invest. Ophthalmol. Vis. Sci.* **38**, S614 (1997).

Acknowledgements. We thank C. Raeburn of the LMB Instrumentation Workshop for assistance. This work was supported by the HFSP M.S.S. held a scholarship of the University of Cambridge M.B. Ph.D. Programme.

Correspondence and requests for materials should be addressed to L.L. (e-mail: lll@mrclmb.cam.ac.uk).

Somatic linker histones cause loss of mesodermal competence in *Xenopus*

Oliver C. Steinbach*, Alan P. Wolffe† & Ralph A.W. Rupp*

* Friedrich Miescher Laboratorium der Max Planck-Gesellschaft, Spemannstrasse 37–39, D-72076 Tübingen, Germany

† Laboratory of Molecular Embryology, National Institute of Child Health and Human Development, NIH, Bethesda, Maryland 20892, USA

In *Xenopus*, cells from the animal hemisphere are competent to form mesodermal tissues from the morula through to the blastula stage¹. Loss of mesodermal competence at early gastrula is programmed cell-autonomously, and occurs even in single cells at the appropriate stage². To determine the mechanism by which this occurs, we have been investigating a concomitant, global change in expression of H1 linker histone subtypes. H1 histones are usually considered to be general repressors of transcription³, but in *Xenopus* they are increasingly thought to have selective functions in transcriptional regulation^{4–6}. *Xenopus* eggs and embryos at stages before the midblastula transition⁷ are deficient

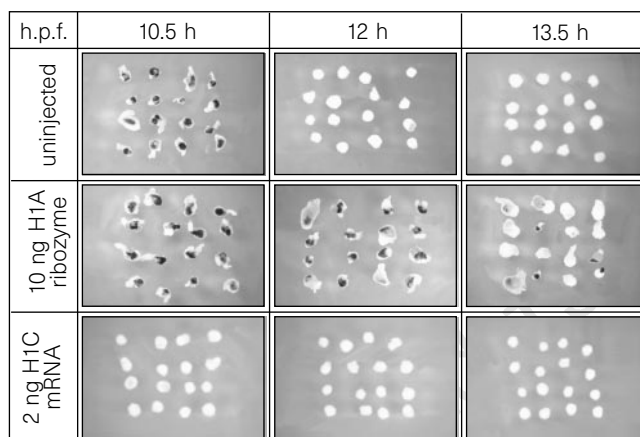


Figure 1 Loss of muscle-forming competence. Animal-cap populations from uninjected, H1A ribozyme- or H1C mRNA-injected embryos were exposed to activin at the indicated time points cultured until tadpole stage (NF 34)¹², and immunostained with the MF20-monoclonal antibody²⁹ for the muscle-differentiation marker myosin heavy chain. At 15 h.p.f. even ribozyme-injected explants were myosin negative (data not shown).

in histone H1 protein, but contain an oocyte-specific variant called histone B4 or H1M. After the midblastula transition, histone B4 is progressively substituted by three somatic histone H1 variants, and replacement is complete by early neurula^{8,9}. Here we report that accumulation of somatic H1 protein is rate limiting for the loss of mesodermal competence. This involves selective transcriptional silencing of regulatory genes required for mesodermal differentiation pathways, like muscle, by somatic, but not maternal, H1 protein.

The induction of mesodermal tissues in animal-cap explants by activin¹⁰, a member of the transforming growth factor (TGF)- β family, provides a well-characterized *in vitro* system to study the loss of mesodermal competence (LMC). We applied activin at 45-min intervals to sibling populations of successively older animal caps, which had been explanted shortly after the midblastula transition. Induction causes stable changes in gene expression patterns, leading ultimately to differentiated mesodermal cell types such as muscle, so this experiment allows us to determine retrospectively whether a cell had been competent to respond to activin at an earlier developmental time, by probing for the expression of marker genes.

Under these conditions, explants ceased both to elongate¹¹ and to form muscle in response to activin between 10.5 and 12 hours post-fertilization (h.p.f.) (Fig. 1), around stage NF 11 (ref. 12). This result was corroborated by a quantitative analysis of mRNA for the myogenic determination factor MyoD¹³, which is normally upregulated by mesoderm induction in the muscle-forming region of the frog embryo^{14,15}. Activin induction of MyoD transcription in animal caps ended at 11.25–12 h.p.f. (Fig. 2a), consistent with the duration of induction of myosin heavy-chain protein. The *brachyury* gene¹⁶, which is expressed throughout the mesoderm, responded in a similar way (Fig. 3a). Taken together, these results are in good agreement with previous data on LMC^{1,2}. Unexpectedly, we found that three other target genes of mesoderm induction displayed a strikingly different activin response profile. The *bmp-4* (ref. 17) and *wnt-8* (ref. 18) genes, which encode signalling factors, were maximally induced at 12–12.75 h.p.f., whereas the locus of the erythroid transcription factor *gata-1* (ref. 19) became inducible only at the end of mesodermal competence (Fig. 3d, g, k). It seems that LMC occurs despite the persistence of a functional activin signalling pathway.

This result questions the prevailing, although unproven, view that LMC is regulated at the level of signal transduction. But LMC may be caused by the loss of transcriptional competence of the regulatory genes required for mesodermal differentiation, in which

case chromatin composition and structure may be important^{20,21}. The rate of accumulation of somatic H1 variants from the mid-blastula transition to the neurula stage has been successfully manipulated^{4,5}. Normally, the H1A subtype constitutes most of the total somatic H1 protein, with the other subtypes H1B and H1C, together making up $\leq 5\%$ (ref. 9). Microinjection of targeted ribozymes into fertilized eggs was used to substantially decrease the accumulation of H1A, or precocious expression of somatic H1 was achieved by H1C mRNA injection. Oocyte 5S RNA transcription at the midblastula transition was then prolonged by H1A ablation, but precociously silenced by H1C overexpression. The transcription of other genes, such as somatic 5S RNA or TFIIIA, was normal under these conditions^{4,5}. The selectivity of these effects led us to use previously characterized molecular reagents⁴ to test whether somatic H1 might also be involved in LMC at gastrula stages.

Injection of the H1A ribozyme severely reduced the somatic accumulation of H1 protein until the neurula stage (see Supplementary information, Fig. S1). Explants derived from such embryos clearly showed a prolonged competence to activin of about 3–4 h (up to neurula¹²), depending on the amount of injected H1A ribozyme; this is indicated by both high-level *MyoD* transcription and the formation of differentiated muscle (Figs 1, 2a). In contrast, injections of the tRNA backbone without H1A ribozyme sequences had no effect (Fig. 2a; data not shown for muscle). Most important, the H1A ribozyme-injected embryos were not delayed in their developmental progression. They were indistinguishable from control siblings with respect to activation of the *GS17* gene²² (Fig. 4a) at the midblastula transition and the temporal appearance of the blastopore pigmentation line at the start of gastrulation¹² (see Supplementary information, Fig. S2). LMC occurred eventually, even in H1A-depleted cells, although it is not known whether this was due to the gradual decay of the injected ribozyme or to phenomena unrelated to H1.

If accumulation of H1A protein is rate-limiting for the loss of muscle-forming competence, then precocious expression of somatic H1 protein should render animal caps refractory to activin at an earlier developmental stage than normal. To test this hypothesis, we injected mRNA encoding the somatic H1C subtype into embryos⁴. Overexpressed H1C protein was incorporated into chromatin at the blastula stage and later (see Supplementary information, Fig. S1), without disturbing the normal chromatin structure⁴. On activin stimulation, H1C-injected explants ceased to form muscle precociously (Fig. 1) and failed to induce *XMyoD* transcription in a dose-dependent manner up to 1.5 h earlier than uninjected explants (Fig. 2b). Control injections of 5 ng β -galactosidase mRNA had no effect on mesodermal competence (Fig. 2b), and developmental progression and the vitality of injected embryos were normal (see Fig. 4a for *GS17* expression). Finally, the fact that the ribozyme specifically targets the endogenous H1A subtype but not the injected H1C mRNA⁴ allowed us to co-inject the two molecules. These experiments demonstrated that the antagonistic phenotypes of H1C overexpression and H1A ablation

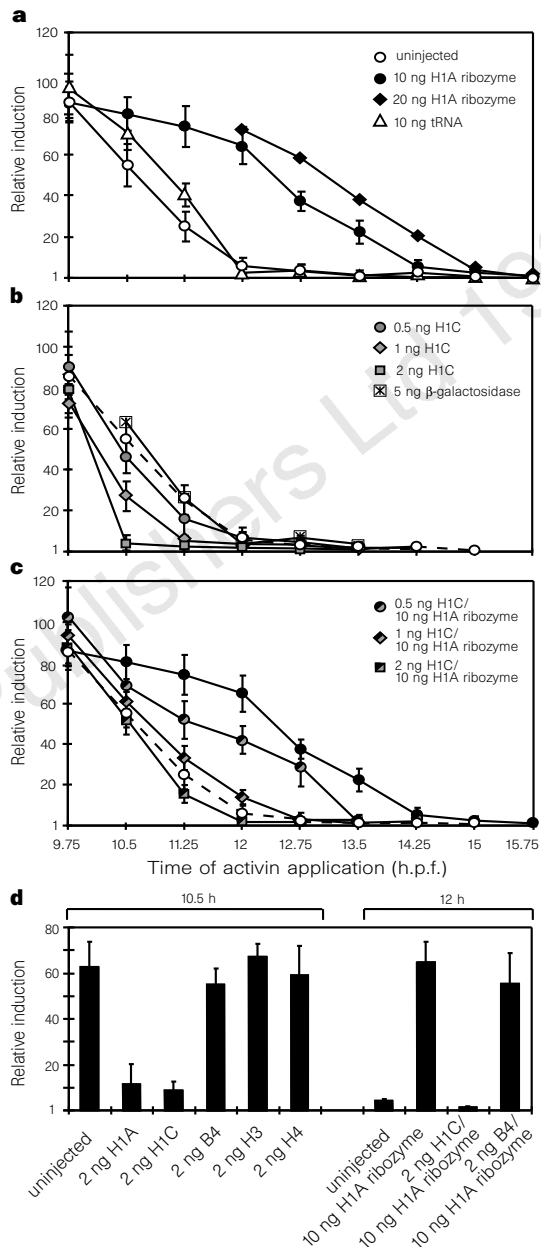


Figure 2 Somatic H1 protein levels control *myoD* induction. Animal caps from embryo populations, which were injected with the RNAs indicated, were analysed for activin induction of the *XmyoD* gene at tailbud stage (error bars, s.d.). **a**, H1A ablation by ribozyme injection. **b**, H1C overexpression by mRNA injection. **c**, Rescue by co-injection of H1A ribozyme and H1C mRNA. In **b** and **c**, the activin response curve of uninjected explants (**a**) is included as reference (broken line). **d**, Overexpression of various linker and core histones.

Table 1 Quantification of H1A ribozyme-dependent phenotypes

RNA injected (ng per embryo)	Target of injection	<i>n</i> (%)	wt (%)	pt (%)
Uninjected	MZ	105 (100)	105 (100)	0 (0)*
tRNA (10)	MZ	65 (78)	65 (100)	0 (0)*
H1A ribozyme (10)	MZ	168 (60)	76 (45)	92 (55)*
Uninjected	NP	115 (91)	115 (100)	0 (0)†
tRNA (10)	NP	61 (92)	61 (100)	0 (0)†
H1A ribozyme (1.2)	NP	118 (78)	110 (93)	8 (7)†
H1A ribozyme (2.4)	NP	36 (69)	34 (94)	2 (6)†
H1A ribozyme (4.8)	NP	114 (72)	106 (93)	8 (7)†
H1A ribozyme (10)	NP	231 (75)	168 (73)	63 (27)†
H1A ribozyme (20)	NP	106 (57)	48 (45)	58 (55)†

MZ, marginal zone injections; NP, neural plate injections; *n*, number (percentage of total embryos in brackets) of neurula embryos (NF 14–16) without gastrulation defects; wt, number (percentage of *n* in brackets) of embryos with wild-type *myoD* expression pattern; pt, number (percentage of *n* in brackets) of embryos with ectopic *myoD* expression pattern.

* Enlarged somitic expression domain.

† Mosaic expression in neural plate.

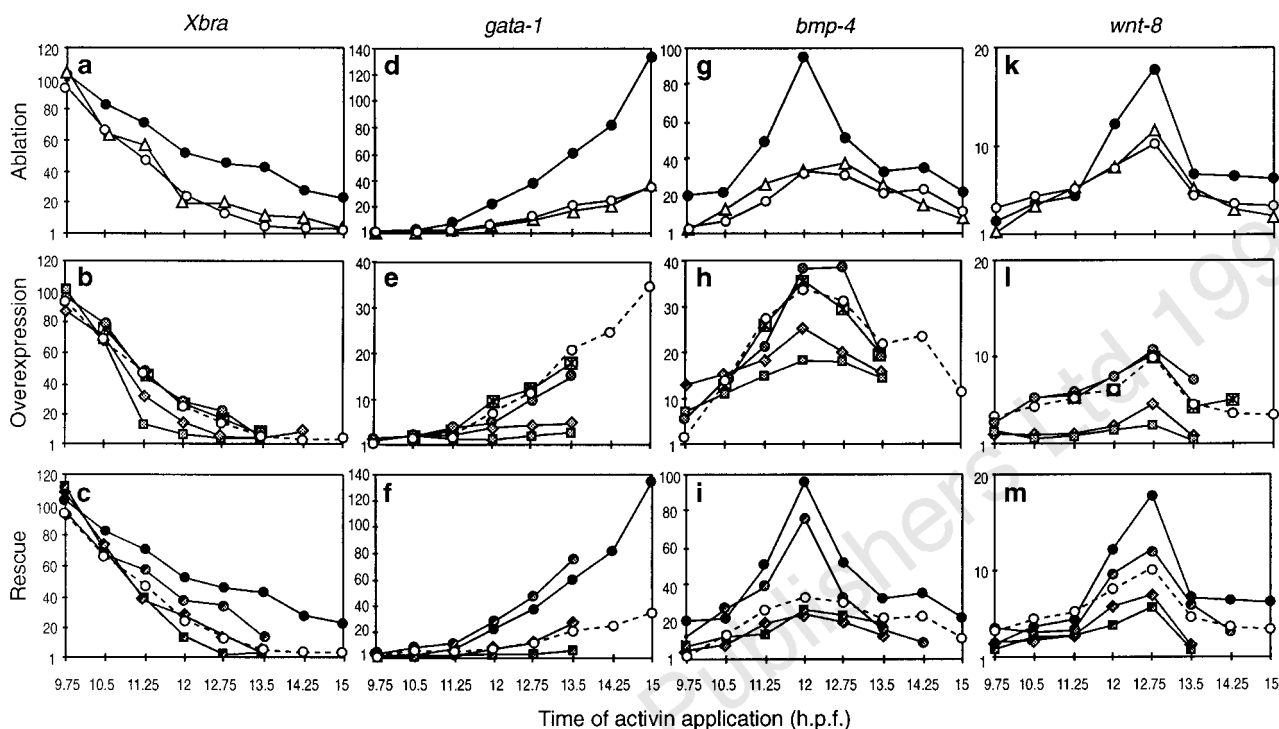


Figure 3 General effects of H1 perturbations for mesoderm induction. RNA samples from experiments in Fig. 2 were reprobbed for expression of other mesodermal loci. Genes are indicated on top, and types of H1 perturbation to the left. Symbols

represent the same RNA combinations as in Fig. 2. Note that y-axes, which show relative induction of mRNA levels, have different scales. Error bars, which were comparable to those for myoD induction (Fig. 2), have been omitted for simplicity.

were indeed due to changes in the amount of somatic linker histone, because the phenotypes were neutralized in a dose-dependent manner (Fig. 2c).

The phenotypes of precocious or extended LMC could either be due to quantitative alterations of total H1 protein levels, or they could involve distinct functions for maternal and somatic linker histone subtypes. To distinguish between these two possibilities, we separately injected mRNAs encoding linker histones H1A, H1C or B4 into animal caps and tested for a reduction in myoD activation brought about by activin (Fig. 2d). Overexpression of either somatic H1A or H1C reduced myoD induction to a similar extent. In contrast, the maternal linker variant B4 had hardly any effect, although exogenous B4 protein was maintained in the chromatin of injected gastrulae (see Supplementary information, Fig. S1). Even when the endogenous somatic histone H1 levels were reduced by H1A ribozyme co-injection, the maternal linker histone could not antagonize the ribozyme phenotype at the time of normal LMC (Fig. 2d). Furthermore, overexpression of core histones H3 and H4 had no effect on MyoD induction (Fig. 2d).

To assess whether the rate-limiting effects of H1 on muscle induction reflect general consequences for mesoderm induction, we reprobbed the same RNA sets for the expression of other genes (Fig. 3). We found that H1A ablation (Fig. 3, top) resulted in all cases in higher mRNA levels upon activin treatment, although temporal characteristics of the individual induction profiles remained. H1C overexpression (Fig. 3, middle) silenced activin-dependent transcription in most cases, the exception being Bmp-4 induction. Finally, coexpression of both RNAs (Fig. 3, bottom) resulted in intermediate phenotypes. We conclude that all tested target genes of mesoderm induction responded to alterations in somatic H1 protein levels in a coherent but gene-specific manner.

These effects do not simply reflect global changes in cellular transcription. First, in the absence of activin, mRNA levels of mesodermal genes were similar in control and H1-manipulated explants (see Supplementary information, Fig. S3). This indicates that the activin requirements for high-level transcription was not

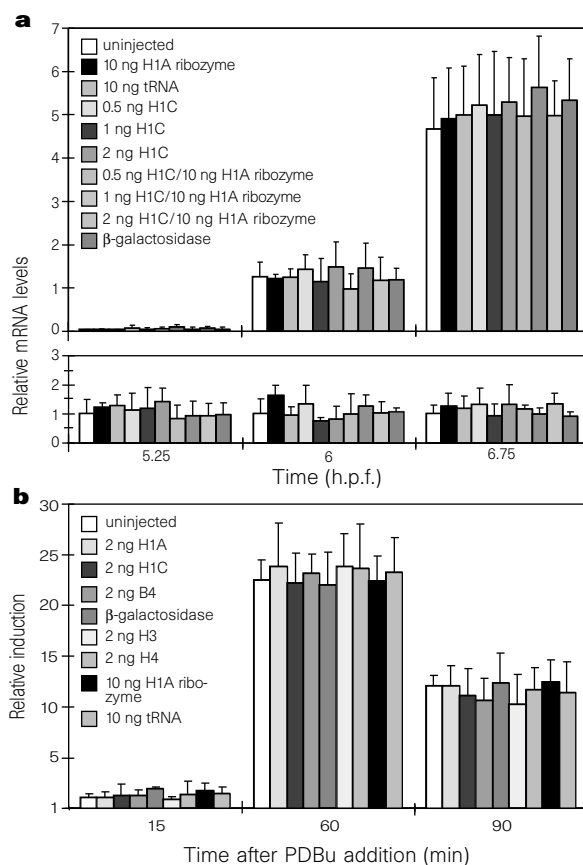


Figure 4 Normal expression of non-mesodermal genes in H1-perturbed chromatin of animal caps. **a**, Relative mRNA levels of genes for GS17 (top) and histone H4 (bottom) at midblastula. **b**, Transient induction of *c-fos* transcription by application of 200 nM phorbol dibutyrate (PDBu) at gastrula (11 h.p.f.). Error bars, s.d.

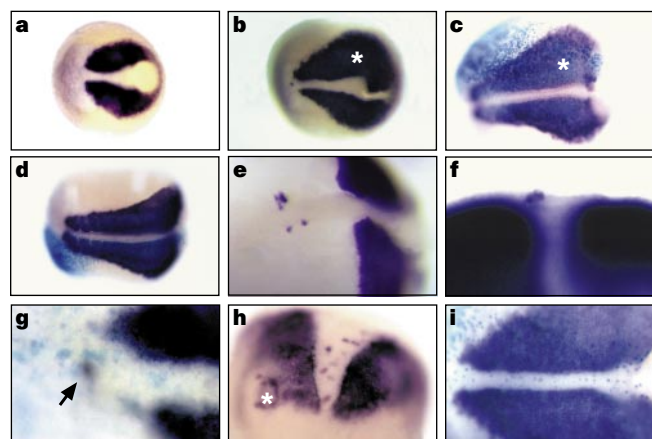


Figure 5 Phenotypic analysis of H1 ablation *in vivo*. *XMyoDb* *in situ* hybridizations at neurula stage. **a**, Uninjected embryo. **a–i**, Injections of 10 ng RNAs were targeted either to the mesoderm (**b–d**) or to the neural plate (**e–i**): H1A ribozyme, **b, c, e–h**; tRNA, **d, i**. In **c, d, g** and **i**, 100 pg β -gal mRNA was co-injected as a lineage label. The views are dorsal, rostral left, except for **f** and **h** (which are dorsotangential views from the anterior).

relieved, and that the chromosomal silencing process was not destabilized in general. Second, mRNA levels of constitutively active (for example, histone H4, Fig. 4a; EF1- α , data not shown) or developmentally regulated genes (GS17, Fig. 4a) did not depend on levels of H1 protein. Finally, transcription of the *Xenopus c-fos* gene can be stimulated by phorbol dibutyrate²³ under all experimental conditions tested (Fig. 4b). In particular, temporal and quantitative aspects of the transient *fos* induction were normal in H1-perturbed chromatin at the developmental time when mesodermal genes were affected. Taken together, these results indicate that there is a regulatory phenomenon of remarkable selectivity.

Are these *in vitro* results relevant for endogenous induction processes? Whole-mount RNA *in situ* hybridization of control embryos revealed the symmetrical MyoD staining pattern (Fig. 5a) characteristic of its expression in the presomitic mesoderm of early neurulae²⁴. Unilateral injection of the H1A ribozyme into the prospective muscle-forming region of the embryo caused an enlargement of the expression domain of MyoD (92 of 168 embryos, 55%) on the injected side (Fig. 5b, c asterisks; Table 1), whereas the injection of tRNA had no effect (Fig. 5d; Table 1). We also investigated whether mesodermal gene expression might occur in the neural plate if mesodermal competence persists. By injecting the A1 and A2 blastomeres of 32-cell embryos, we delivered the injected tRNA or ribozyme to the neural plate²⁵. Depending on the H1A ribozyme dose, up to 55% of the injected embryos (Table 1) showed ectopic MyoD expression in small clusters or single cells in the neural plate (Fig. 5e, f). Lineage tracing with co-injected β -gal mRNA suggested that the effect of the H1A ribozyme is cell autonomous, and that ectopic induction is mosaic (Fig. 5g, arrow). A few embryos displayed additional MyoD staining in deep layers, directly adjacent to the myotome (Fig. 5h, asterisk). Similarly, *Xbra* was ectopically activated under these conditions, although at lower frequency (13 of 159, 8%; see Supplementary information, Fig. S4). In contrast, we found no evidence for a leaky induction of *myoD* among more than 240 control or tRNA-injected embryos (Fig. 5a, i; Table 1). These results suggest that somatic H1A protein has at least two new functions in normal development: to restrict the number of muscle precursors, and to uncouple some redundancy between mesoderm and neural induction processes.

The temporal restriction of the H1 transition up to neurula stage⁹ implies that there are H1-independent mechanisms that control competence at later stages. For the early period of *Xenopus* development, our results, which demonstrate specific effects both with respect to H1 subtypes and the genes affected, indicate strongly that

H1 has selective functions in transcriptional regulation. Genetic data from tetrahymena, yeast and mice (reviewed in ref. 21) support this view. In the case of LMC, H1 may either silence the expression of a critical component(s) of the activin pathway, or render mesodermal genes insensitive to activin stimulation. By using chromatin-mapping experiments, we hope to identify the target genes of H1 proteins, which will allow us to address the functions of H1 in more detail. Our results so far have provided both a mechanism for the long-standing phenomenon of LMC and an example of general chromatin components being involved in pattern-formation processes during vertebrate development. □

Methods

General embryological techniques, *in situ* hybridization with digoxigenin-labelled antisense probes, and immunocytochemistry for myosin heavy-chain protein have been described^{26,27}. Activin A was applied as a 4-fold dilution of conditioned medium from P388D1 cells²⁸ to animal caps, which were mechanically reopened to allow growth-factor access. The H1A ribozyme, tyrosine tRNA, and the H1C, H3 and H4 cDNAs have been described and functionally characterized in detail⁴. The coding region of H1A has been cloned by the polymerase chain reaction (PCR) into the same vector (pSP64polyA, Promega) and was verified by sequencing (O.C.S., unpublished data). Capped transcripts were synthesized as described²⁶. Synthetic histone mRNAs were translated with comparable efficiency *in vitro* (data not shown). A detailed description of the RT-PCR method, including primer sequences, is available from R.A.W.R. on request (see also ref. 15). Our conditions ensure a direct correlation between RNA template abundance and PCR product amounts and strict dependence of PCR products on cDNA synthesis. Data points were calculated as means of duplicate samples (3 animal caps each) from at least 3 independent experiments. They represent increases in steady-state mRNA levels of induced over uninduced explants after normalization to histone H4 mRNA levels.

Received 19 May; accepted 4 July 1997.

- Jones, E. A. & Woodland, H. R. The development of animal cap cells in *Xenopus*: a measure of the start of animal cap competence to form mesoderm. *Development* **101**, 557–563 (1987).
- Grainger, R. M. & Gurdon, J. B. Loss of competence in amphibian induction can take place in single nondividing cells. *Proc. Natl Acad. Sci. USA* **86**, 1900–1904 (1989).
- Paranjape, S. M., Kamakaka, R. T. & Kadonaga, J. T. Role of chromatin structure in the regulation of transcription by RNA polymerase II. *Annu. Rev. Biochem.* **63**, 265–297 (1994).
- Bouvet, P., Dimitrov, S. & Wolffe, A. P. Specific regulation of *Xenopus* chromosomal 5S rRNA gene transcription *in vivo* by histone H1. *Genes Dev.* **8**, 1147–1159 (1994).
- Kandolf, H. The H1A histone variant is an *in vivo* repressor of oocyte-type 5S gene transcription in *Xenopus laevis* embryos. *Proc. Natl Acad. Sci. USA* **91**, 7257–7260 (1994).
- Patterson, D. & Wolffe, A. P. Developmental roles for chromatin and chromosomal structure. *Dev. Biol.* **173**, 2–13 (1996).
- Newport, J. & Kirschner, M. A major developmental transition in early *Xenopus* embryos: II. control of the onset of transcription. *Cell* **30**, 687–696 (1982).
- Dworkin-Rastl, E., Kandolf, H. & Smith, R. C. The maternal histone H1 variant, H1M (B4 protein), is the predominant H1 histone in *Xenopus* pregastrula embryos. *Dev. Biol.* **161**, 425–439 (1994).
- Dimitrov, S., Almouzni, G., Dasso, M. & Wolffe, A. P. Chromatin transitions during early *Xenopus* embryogenesis: changes in histone H4 acetylation and in linker histone type. *Dev. Biol.* **160**, 214–227 (1993).
- Smith, J. C. Mesoderm-inducing factors and mesodermal patterning. *Curr. Opin. Cell Biol.* **7**, 856–861 (1995).
- Symes, K. & Smith, J. C. Gastrulation movements provide an early marker of mesoderm induction in *Xenopus laevis*. *Development* **101**, 339–349 (1987).
- Nieuwkoop, P. D. & Faber, J. *Normal table of Xenopus laevis* (Daudin) (North-Holland, Amsterdam, 1967).
- Rudnicki, M. A. *et al.* MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* **75**, 1351–1359 (1993).
- Hopwood, N. D., Pluck, A. & Gurdon, J. B. MyoD expression in the forming somites is an early response to mesoderm induction in *Xenopus* embryos. *EMBO J.* **8**, 3409–3417 (1989).
- Rupp, R. A. W. & Weintraub, H. Ubiquitous MyoD transcription at the midblastula transition precedes induction-dependent MyoD expression in presumptive mesoderm of *X. laevis*. *Cell* **65**, 927–937 (1991).
- Smith, J. C., Price, B. M. J., Green, J. B. A., Weigel, D. & Hermann, B. G. Expression of a *Xenopus* homolog of *Brachyury* (T) is an immediate-early response to mesoderm induction. *Cell* **67**, 79–87 (1991).
- Köster, M. *et al.* Bone morphogenetic protein 4 (BMP-4), a member of the TGF- β family, in early embryos of *Xenopus laevis*: analysis of mesoderm inducing activity. *Mech. Dev.* **33**, 191–200 (1991).
- Christian, J. L., McMahon, J. A., McMahon, A. P. & Moon, R. T. *Xwnt-8*, a *Xenopus Wnt-1/int-1*-related gene responsive to mesoderm-inducing growth factors, may play a role in ventral mesodermal patterning during embryogenesis. *Development* **111**, 1045–1055 (1991).
- Pevny, L. *et al.* Erythroid differentiation in chimeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257–260 (1991).
- Svaren, J. & Hörz, W. Regulation of gene expression by nucleosomes. *Curr. Opin. Genet. Dev.* **6**, 164–170 (1996).
- Hartog, G. A. & Winston, F. Nucleosomes and transcription: recent lessons from genetics. *Curr. Opin. Genet. Dev.* **7**, 192–198 (1997).
- Krieg, P. & Melton, D. A. Developmental regulation of a gastrula-specific gene injected into fertilized *Xenopus* eggs. *EMBO J.* **4**, 3463–3471 (1985).
- Tsuda, T., Hamamori, Y., Yamashita, T., Fukumoto, Y. & Takai, Y. Involvement of three intracellular

- messenger systems, protein kinase C, calcium ion and cyclic AMP, in the regulation of *c-fos* gene expression in Swiss 3T3 cells. *FEBS Lett.* **208**, 39–42 (1986).
24. Frank, D. & Harland, R. M. Transient expression of XMyoD in non-somitic mesoderm of *Xenopus* gastrulae. *Development* **113**, 1387–1393 (1991).
25. Dale, L. & Slack, J. M. W. Fate map for the 32-cell stage of *Xenopus laevis*. *Development* **99**, 527–551 (1987).
26. Rupp, R. A. W., Snider, L. & Weintraub, H. *Xenopus* embryos regulate the nuclear localization of XMyoD. *Genes Dev.* **8**, 1311–1323 (1994).
27. Steinbeisser, H., Fainsod, A., Niehrs, C., Sasai, Y. & De Robertis, E. M. The role of *gsc* and BMP-4 in dorsal-ventral patterning of the marginal zone in *Xenopus*: a loss-of-function study using antisense RNA. *EMBO J.* **14**, 5230–5243 (1995).
28. Sokol, S., Wong, G. G. & Melton, D. A. A mouse macrophage factor induces head structures and organizes a body axis in *Xenopus*. *Science* **249**, 561–564 (1990).
29. Bader, D., Masaki, T. & Fischman, D. A. Immunohistochemical analysis of myosin heavy chain during avian myogenesis *in vivo* and *in vitro*. *J. Cell Biol.* **95**, 763–770 (1982).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

Acknowledgements. We thank M. Leu for help with RT-PCR; S. Dimitrov and R. Smith for B4/H1 antisera; S. Wendler for PCR primers; and C. Dreyer and H. Steinbeisser for comments.

Correspondence and requests for materials should be addressed to R.A.W.R. (e-mail: rupp@gen.mpiib-tuebingen.mpg.de).

Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA

Jean-Christophe Dantoni, Kanneganti G. K. Murthy*, James L. Manley* & Laszlo Tora

Institut de Génétique et de Biologie Moléculaire et Cellulaire CNRS/INSERM/ULP, Collège de France, BP 163, F-67404 Illkirch Cedex, CU de Strasbourg, France

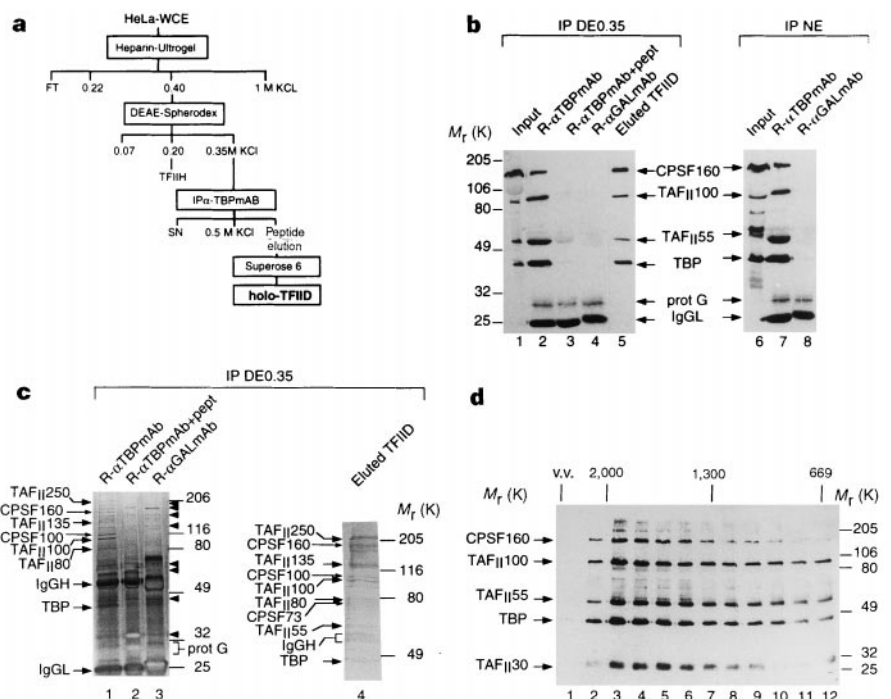
* Department of Biological Sciences, Columbia University, New York, New York 10027, USA

Initiation of transcription by RNA polymerase II from a promoter region on DNA requires the assembly of several initiation factors to form a preinitiation complex. Assembly of this complex^{1,2} is initiated by the binding of the transcription factor TFIID, composed of the TATA-box binding protein (TBP) and TBP-associated factors (TAF_{II}s), to the promoter. We have now characterized an immunopurified TFIID complex which we unexpectedly find

contains the cleavage–polyadenylation specificity factor (CPSF), one of the factors required for formation of the 3' end of messenger RNA^{3,4}. CPSF is brought to the preinitiation complex by TFIID, but after transcription starts, CPSF dissociates from TFIID and becomes associated with the elongating polymerase. We also show that overexpression of recombinant TBP in HeLa cells decreases polyadenylation without affecting the correct initiation of transcription of the reporter gene. This indicates that, owing to incomplete assembly of TFIID on recombinant TBP, CPSF is not brought to the promoter and therefore polyadenylation becomes less efficient. Our observations have thus revealed a link between transcription initiation and elongation by RNA polymerase II and processing of the 3' end of mRNA.

We have previously shown that there are functionally distinct TFIID complexes composed of both common and specific TAF_{II}s (refs 5–7). To find and characterize other TFIID-associated proteins, we carried out a large-scale immunoprecipitation (Fig. 1a, and refs 5, 8) with an anti-TBP monoclonal antibody⁶. Immunoprecipitated TFIID complexes were eluted with an excess of epitope peptide⁶ (Fig. 1c, lane 4) and all the protein species present in this TFIID complex were identified by microsequencing. We found that most of the tryptic peptides corresponded to the previously identified TAF_{II}s, but the sequences of peptides (see Methods) derived from three protein species (of relative molecular masses 150K, 105K and 70K) were different and did not correspond to any known TAF_{II}s. Comparison of the amino-acid sequences of these peptides with a protein database identified them as the 160K, 100K and 73K subunits of the human cleavage–polyadenylation specificity factor CPSF^{9–11}. CPSF is a complex that plays a role in 3' processing of precursor mRNAs by binding to the AAUAAA signal sequence and coordinating interactions with other polyadenylation factors^{3,4}. The presence of CPSF in our immunopurified and eluted TFIID (hereafter called holo-TFIID) was also demonstrated by western blot analysis using an antibody raised against the 160K subunit of CPSF (CPSF-160) (Fig. 1b). In control immunoprecipitations, carried out with either an unrelated antibody (anti-Gal4 monoclonal antibody) or with epitope-peptide-saturated anti-TBP antibody, no CPSF-160 was detected (Fig. 1b, lanes 3, 4, and Fig. 1c, lanes 2, 3). Moreover, TFIID and CPSF were also co-immunoprecipitated from unfractionated nuclear extracts by the anti-TBP antibody (Fig. 1b, lane 7)

Figure 1 CPSF is a component of holo-TFIID. **a**, Chromatography protocol used to purify TFIID. **b**, The DE0.35 fraction (lane 1; and see **a**) or a HeLa cell nuclear extract (NE) (lane 6) were immunoprecipitated (IP) with anti-TBP mAb, the epitope-peptide-saturated anti-TBP mAb, or an unrelated mAb bound to protein-G resin (R). Input fractions (lanes 1 and 6), resin-bound proteins and peptide-eluted TFIID were analysed by immunoblotting. **c**, Fractions from the DE0.35 IP were analysed by silver staining. Non-specific proteins are indicated by arrowheads. The identity of each indicated TFIID and CPSF subunit, present in the large-scale TFIID preparation, was determined by peptide sequencing (lane 4). **d**, Fractions eluting from the gel-filtration column were analysed by western blotting. The void volume (v.v.) and the elution peaks of relative molecular mass standards are indicated.



but not by the unrelated antibody (Fig. 1b, lane 8), indicating that CPSF associates with TFIID under physiological conditions; CPSF can, however, be dissociated from TFIID by 1 M KCl. Taken together, these results indicate that there is a TFIID population associated with the intact CPSF complex in HeLa cells.

To analyse this association of CPSF with the TFIID complex, holo-TFIID was further purified by Superose-6 gel-filtration chromatography. Analysis of the gel-filtration chromatographic fractions by either western blotting (Fig. 1d) or silver staining (data not shown) of an SDS-polyacrylamide gel indicated that CPSF coelutes with the TAF_{II}s and TBP in a single peak, with an apparent relative molecular mass of 1,300K–2,000K. No CPSF was found in the 300K–500K range (data not shown) where free CPSF complex would be expected to elute^{12,13}. These results indicate that the isolated holo-TFIID complex is stably associated with CPSF.

To identify the subunits of TFIID that interact with CPSF, the complementary DNAs encoding human TAF_{II}s and TBP were inserted in baculovirus expression vectors^{7,9,14} and expressed individually together with His-CPSF-160 in Sf9 cells. Nuclear extracts were prepared and protein expression was tested by western blot analysis (Fig. 2a–d, lanes 1 and 3). From these extracts, either His-CPSF-160 was retained on Ni-NTA resin (Fig. 2a, b) or the TAF_{II}s were immunoprecipitated with the appropriate monoclonal antibodies (Fig. 2c, d) and the bound proteins analysed by western blotting (Fig. 2a–d). Extracts in which neither CPSF-160 nor the TAF_{II}s were coexpressed served as negative controls for Ni-NTA chromatography and the immunoprecipitation, respectively. These results indicate that CPSF-160 binds strongly to hTAF_{II}100, hTAF_{II}55 and hTAF_{II}20, weakly to hTAF_{II}28 and hTAF_{II}18, and not at all to hTAF_{II}250, hTAF_{II}68 or hTAF_{II}30 (Fig. 2e), suggesting that CPSF-160 may contact one or more of these hTAF_{II}s in the TFIID complex.

To investigate the relevance of the TFIID–CPSF interaction during assembly of the preinitiation complex (PIC) and subsequent

transcription initiation and elongation, immunoprecipitations were carried out with an anti-TBP monoclonal antibody and another raised against the C-terminal domain (CTD) of the largest subunit of RNA polymerase II (Pol II) during different stages of transcription *in vitro*. In this reconstituted system, the general Pol II transcription factors were either recombinant proteins (TFIIA α/β , TFIIB, TFIIE α , TFIIE β , TFIIF α and TFIIF β) or endogenous factors (TFIIH and Pol II)^{8,14,15}. None of these fractions contained CPSF (Fig. 3a, lane 1), the only source of which was the immunopurified holo-TFIID (Fig. 3a, lane 2). This system supports transcription from either the supercoiled DNA template containing the minimal adenovirus major late promoter (AdMLP)¹⁴ or the supercoiled TI template containing the TATA box and initiator of AdMLP¹⁶ (data not shown). On the latter template, elongation complexes will stall in the absence of GTP, enabling elongation complexes to be isolated. After PIC assembly, the anti-CTD antibody co-immunoprecipitated not only TBP, but also CPSF-160 (Fig. 3a, lane 7), together with Pol II. About 25% of the input TFIID and almost 50% of the input CPSF were stably associated with the PIC (Fig. 3a, lanes 7 and 8). After nucleotide addition, when Pol II enters the elongation phase and the CTD of the largest Pol II subunit becomes highly phosphorylated¹⁷, the anti-CTD antibody could no longer co-immunoprecipitate TBP (or TFIID) with Pol II (now hyperphosphorylated: Pol IIO, lane 10 in Fig. 3a), confirming that Pol II had left the PIC. CPSF still co-immunoprecipitated Pol II, indicating that it had become associated with the elongating polymerase (Fig. 3a, lane 10). In agreement with this finding, after initiation the anti-TBP antibody immunoprecipitated only trace amounts of CPSF (Fig. 3a, lane 11). Note that when the DNA template was omitted, anti-CTD antibody did not immunoprecipitate either CPSF or TBP (Fig. 3a, lane 4). These results show that although CPSF is stably associated with the PIC, after transcription initiation it dissociates from the TFIID complex and associates with the Pol II elongation complex.

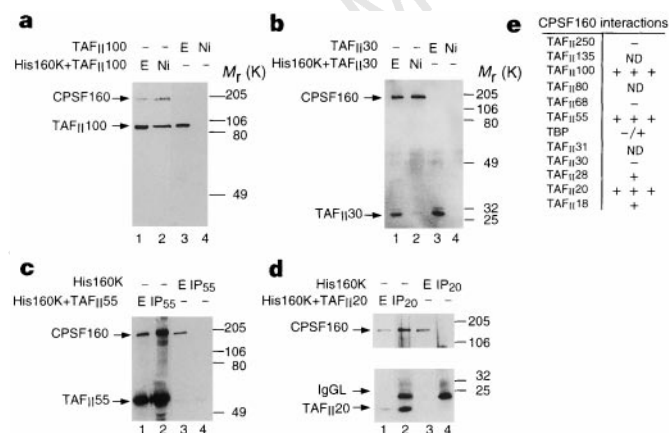


Figure 2 Interactions of the CPSF-160 with individual components of the TFIID complex. Sf9 cells were co-infected with baculoviruses expressing His-tagged CPSF-160 (His160K), hTAF_{II}s and TBP. Nuclear extracts (E) were made and either His160K was retained on Ni-NTA resin (**a**, **b**) or the TAF_{II}s were immunoprecipitated with the appropriate mAbs (**c**, **d**). Bound proteins were then analysed by western blotting. TAF_{II}s and His160K were also expressed individually as controls. IgGL, immunoglobulin light chain. **e**, Summary of interactions of CPSF-160 with individual components of the TFIID complex. ND, not determined.

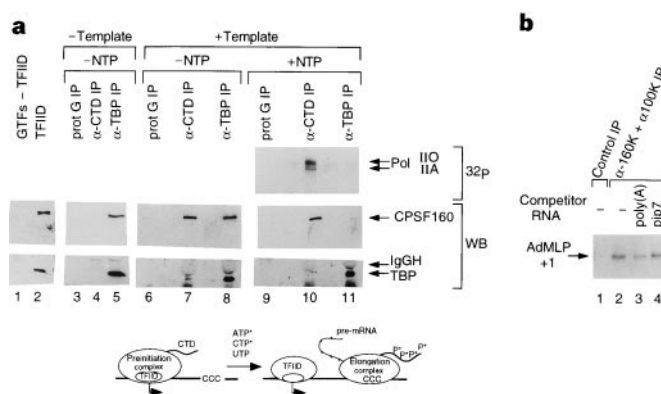


Figure 3 Fate of CPSF during PIC assembly and transcription. **a**, Reactions were assembled with purified general transcription factors (GTFs), with or without the template, and immunoprecipitated with the indicated mAbs (or protein-G resin alone (prot G)) either after PIC formation (–NTP) or following initiation (+NTP). Bound proteins were analysed by western blot (WB) with mAbs raised against TBP and CPSF-160. Input factors (GTFs-TFIID) and holo-TFIID alone were tested by immunoblotting. The hyperphosphorylation of Pol II (Pol IIO) was also monitored (³²P). IgGH: immunoglobulin heavy chain. The model underneath depicts the preinitiation complex and, after nucleotide addition, the stalled elongation complex on the G-less cassette containing template. **b**, CPSF present in the elongation complex can associate with the newly synthesized pre-mRNA. Transcriptions were assembled, using the AdMLP template, and reactions were then immunoprecipitated with a pre-immune serum (Control IP) or a mixture of anti-CPSF-100 and -CPSF-160 sera (α-160K + α-100K IP). CPSF-bound transcripts were analysed by S1 nuclease mapping. Where indicated, 200 ng competitor RNA, containing either a poly(A) signal or not (pip7), were added to the reactions before the IPs.

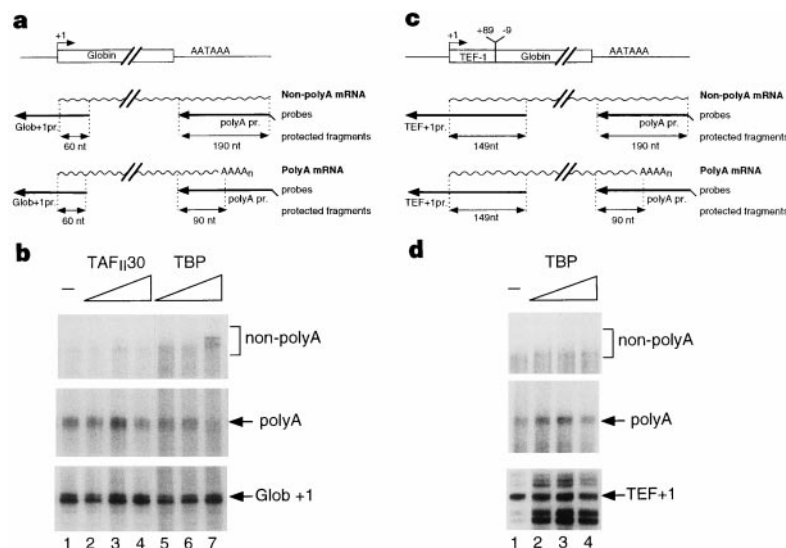


Figure 4 Overexpressed TBP inhibits polyadenylation of transcripts initiated from the TATA-box-containing promoter of the β -globin gene. **a, c**, Structure of the reporter genes and scheme of S1 nuclease mapping. Probes (pr) and the length of expected products (in nucleotides (nt)) after S1 nuclease digestion are shown. The +1 probes (Glob+1 pr. and TEF+1 pr.) map the correctly initiated transcripts from the globin and the TEF-1 promoters. HeLa cells were co-transfected with the **b**, globin, or **d**, TEF-1 reporter genes, together with expression vectors for either human (h)TAF_{II}30 or hTBP. Quantitative S1 nuclease analysis was carried out using 10–15 μ g RNA.

Next we tested whether CPSF present in the transcription elongation complex could associate with newly synthesized pre-mRNA. Transcription reactions were assembled on the AdMLP template and following transcription initiation, immunoprecipitations were carried out with either a mixture of anti-CPSF-160 and anti-CPSF-100 antisera or a preimmune serum. Any bound pre-mRNA was purified and subjected to S1 nuclease protection analysis¹⁸. The anti-CPSF-160 and anti-CPSF-100 antibodies together (Fig. 3b, lane 2), but not the control serum (Fig. 3b, lane 1), immunoprecipitated the newly synthesized pre-mRNA. The RNA transcribed from the β -globin gene contains a perfect AAUAAA sequence. To determine whether this is essential for CPSF binding, competitor RNAs with or without the adenovirus L3 poly(A) site⁹ were added to the transcription reactions before immunoprecipitation. RNA containing the poly(A) site (Fig. 3b, lane 3), but not the control RNA (Fig. 3b, lane 4), reduced the amount of immunoprecipitated correctly initiated RNA 2–3-fold. This indicates that CPSF, brought to the PIC by TFIID, is functional and can interact with the poly(A) site of the new β -globin transcript. Thus CPSF is probably directly transferred from the PIC to Pol II, and then to the nascent RNA.

To put these findings into a physiological context, transient transfection experiments were carried out in HeLa cells with increasing amounts of vectors expressing human TBP or TAF_{II}30, together with either the rabbit β -globin gene sequences (including the TATA-box-containing promoter, the 3' non-translated region, and the poly(A) site (Fig. 4a)), or a reporter in which the globin gene sequences are under the control of the TATA-less promoter of the TEF-1 gene¹⁹ (Fig. 4c). Transfection of increasing amounts of human TBP expression vector resulted in a 3–4-fold decrease in correctly polyadenylated β -globin mRNA and an increase in non-polyadenylated mRNA (Fig. 4b), whereas the total amount of correctly initiated transcript did not change. In contrast, overexpression of TAF_{II}30 had no effect, or only a weak effect, on polyadenylation (Fig. 4b). Moreover, overexpression of TBP did not decrease the amount of polyadenylated transcript from the TATA-less TEF-1 promoter compared with the number of correctly initiated transcripts (Fig. 4d), in agreement with the finding that overexpressed TBP cannot interact with TATA-less promoters *in vivo*²⁰. These results together suggest that overexpressed hTBP competes with the endogenous CPSF-containing holo-TFIID complex for binding to the TATA box-containing promoter, which in turn results in decreased polyadenylation as CPSF is not directly transferred to the elongating polymerase so there is no transcription-coupled recognition of the polyadenylation signal. It appears that the CPSF-containing holo-TFIID complex is a biological entity.

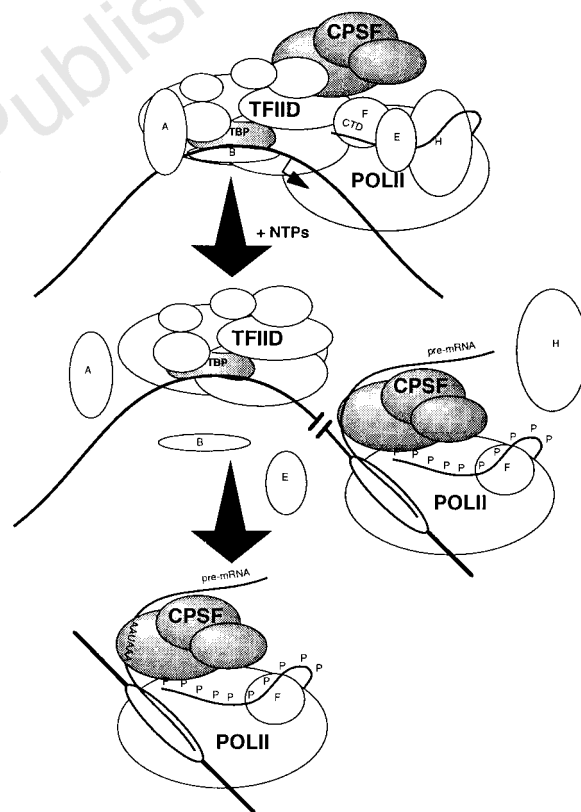


Figure 5 Model depicting the fate of CPSF during transcription initiation, elongation and 3' pre-mRNA processing. CPSF is recruited to the PIC together with TFIID. Upon addition of nucleotides, CPSF dissociates from TFIID and becomes associated with the elongating Pol II. Our model implies that CPSF dissociates from the elongation complex and directs formation of the 3' end. This transition may play a role in the transcription termination process^{28–30}. TFIID transcription factors are represented by A, B, E, F, H; Ps indicate hyperphosphorylation of the carboxy-terminal domain of Pol II.

Our results show that transcription and 3' mRNA processing are mechanistically coupled. TFIID nucleates transcription initiation by Pol II and also initiates polyadenylation by recruiting the CPSF complex to an active promoter, and subsequently, through the polymerase, to the poly(A) site of the nascent RNA (Fig. 5). It remains to be determined whether this transcription initiation-coupled polyadenylation reaction occurs at most class II genes or only at a subset. The fact that CPSF is transferred to Pol II and is

carried with it during elongation is in keeping with the direct interaction reported previously between CPSF and the CTD of Pol II present in the holopolymerase complex²¹. Our results indicate that the hyperphosphorylated form of the CTD is involved in the interaction with CPSF; upon reaching the polyadenylation signal, CPSF probably dissociates from the elongation complex to help define the polyadenylation site and catalyse 3' end formation. □

Methods

Immunoprecipitation and microsequencing. Nuclear extract and the fraction eluting from a DEAE-Spheroex column between 0.2 and 0.35 M KCl (DE0.35 fraction; Fig. 1) were dialysed against 100 mM KCl containing immunoprecipitation (IP) buffer⁶ and immunoprecipitated with the anti-TBP monoclonal antibody (mAb) 2C1 (ref. 6) or with the control anti-Gal4 mAb. For the large-scale immunoprecipitation of TFIID complexes, 100 ml (~100 mg) of the DE0.35 fraction was used with 1 ml protein G-Sepharose and 1 mg mAb 2C1. Bound TFIID was washed with 500 mM KCl containing IP buffer and eluted by using a 1,000-fold excess of the corresponding epitope peptide. All protein species from the eluted TFIID were microsequenced as described^{6,22}. Microsequences obtained from a ~150K polypeptide corresponding to the hCPSF-160 were the following: (708)-SGPEAEGLGSET-(719), (844)-PHLLVHDQELLYEAF-(860), (1379)-NVLD GELLNR-(1388); microsequences from a ~105K polypeptide corresponding to the hCPSF-100 were the following: (471)-IKWDEYGEIHKPEDFLVPELQA-(492), (642)-VDTGVILE-(649) and (1379)-NVLDGELLNR-(1388); a microsequence from a ~70K polypeptide corresponding to the hCPSF-73 was the following: (534)-LTGFVEELIEIQEPAL-(548). The numbers in parentheses indicate the amino acids in the human CPSF-160 (ref. 9), in the bovine CPSF-100 (ref. 12), and in the bovine CPSF-73 (ref. 11) subunits, respectively.

Gel filtration. Holo-TFIID (30 µg) was injected using the Smart System (Pharmacia) onto a Superose-6 gel-filtration column equilibrated in PBS with 0.1% Nonidet P-40. Molecular size markers were separated under the same conditions.

Protein-protein interactions. Baculovirus expression vectors for TAF_{II}58, CPSF-160, Sf9 cell infection and protein expression have been described^{7,9,14}. Nuclear extract from Sf9 cells was prepared as follows: cells were washed with ice-cold PBS containing 20% glycerol and resuspended in IP buffer⁶ containing 100 mM KCl. Cells were lysed by 3 cycles of freeze-thawing and centrifuged. The pellet was resuspended in IP buffer (100 mM KCl), sonicated, and insoluble material removed by centrifugation. Immunoprecipitations with anti-TAF_{II}55, anti-TAF_{II}28, anti-TAF_{II}20 and anti-TAF_{II}18 mAbs were done as described²³⁻²⁵. Purification with Ni²⁺-NTA resin was according to ref. 6, except that the Ni²⁺-NTA resin was washed with IP buffer containing 500 mM KCl.

In vitro transcription-coupled immunoprecipitations. Transcription reactions were assembled using highly purified general Pol II transcription factors as described^{5,14}, except for the experiments shown in Fig. 3a, in which the TI template was used that contained the TATA-box and initiator of AdMLP upstream of a G-less cassette¹⁶. Reactions were incubated for 30 min at 25 °C to allow PIC formation, then were either immunoprecipitated or nucleotides were added and incubation continued for a further 45 min at 25 °C. When the G-less cassette containing template was used, GTP was omitted from the nucleotide mix and transcripts were analysed as described¹⁶. The hyperphosphorylation of Pol II (Pol IIO) was monitored by including 2.5 pmol [γ -³²P]ATP into the NTP mix. Immunoprecipitations were carried out by adding 10 µl protein-G-Sepharose, protein-G-Sepharose-bound anti-CTD mAb (7G5)²⁶, or protein-G-Sepharose-bound anti-TBP mAb (3G3)⁵, in 50 µl IP buffer (100 mM in KCl) and incubating for 2 h at 4 °C. Beads were then extensively washed with IP buffer (200 mM in KCl) and bound proteins were resolved by SDS-PAGE and analysed by western blot. Transcription reactions were stopped by addition of 0.4 µg tRNA in 300 µl IP buffer⁶ and mixed for 1 h at room temperature with 20 µl protein-G-Sepharose that had been pre-incubated either with 20 µl preimmune sera or a mixture of anti-CPSF-100 and anti-CPSF-160 antisera; the beads were then extensively washed with IP buffer containing 200 mM KCl. Bound transcripts were extracted with phenol/chloroform and analysed by quantitative S1 nuclease mapping¹⁸. Competitor RNAs were synthesized as described⁹.

Cell transfections and RNA preparation. HeLa cells were transfected with 5 µg pG1 (ref. 27) or 10 µg pTΔ-530 (ref. 19) reporter plasmids alone or together with 1, 3 or 9 µg expression vector for hTBP²³ and hTAF_{II}30 (ref. 6) as described¹⁸. Cells were collected after 48 h and total RNA prepared. The antisense Glob + 1 and the TEF + 1 probes have been described^{18,19}; the probe overlapping the rabbit β-globin poly(A) site was synthesized by PCR from +1,396 to the BglII site at +1,200 (relative to the transcription start site). RNA was analysed by quantitative S1 nuclease mapping¹⁸.

Received 27 March; accepted 20 June 1997.

- Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327-336 (1996).
- Koleske, A. J. & Young, R. A. The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem. Sci.* **20**, 113-116 (1995).
- Keller, W. No end yet to messenger RNA 3' processing. *Cell* **81**, 829-832 (1995).
- Manley, J. L. A complex protein assembly catalyzes polyadenylation of mRNA precursors. *Curr. Opin. Genet. Dev.* **5**, 222-228 (1995).
- Brou, C. *et al.* Distinct TFIID complexes mediate the effect of different transcriptional activators. *EMBO J.* **12**, 489-499 (1993).
- Jacq, X. *et al.* Human TAF_{II}30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* **79**, 107-117 (1994).
- Bertolotti, A., Lutz, Y., Heard, D. J., Chambon, P. & Tora, L. hTAF_{II}68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J.* **15**, 5022-5031 (1996).
- Gerard, M. *et al.* Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. *J. Biol. Chem.* **266**, 20940-20945 (1991).
- Murthy, K. G. & Manley, J. L. The 160-kD subunit of human cleavage-polyadenylation specificity factor coordinates pre-mRNA 3'-end formation. *Genes Dev.* **9**, 2672-2683 (1995).
- Bienroth, S., Wahle, E., Suter-Crazzolara, C. & Keller, W. Purification of the cleavage and polyadenylation factor involved in the 3'-processing of messenger RNA precursors. *J. Biol. Chem.* **266**, 768-776 (1991).
- Jenny, A., Minvielle-Sebastia, L., Preker, P. J. & Keller, W. Sequence similarity between the 73-kilodalton protein of mammalian CPSF and a subunit of yeast polyadenylation factor I. *Science* **274**, 1514-1517 (1996).
- Jenny, A., Hauri, H. P. & Keller, W. Characterization of cleavage and polyadenylation specificity factor and cloning of its 100-kilodalton subunit. *Mol. Cell. Biol.* **14**, 8183-8190 (1994).
- Murthy, K. G. & Manley, J. L. Characterization of the multisubunit cleavage-polyadenylation specificity factor from calf thymus. *J. Biol. Chem.* **267**, 14804-14811 (1992).
- Dubrovskaya, V. *et al.* Distinct domains of hTAF_{II}100 are required for functional interaction with transcription factor TFIIB (RAP30) and incorporation into the TFIID complex. *EMBO J.* **15**, 3702-3712 (1996).
- Sun, X., Ma, D., Sheldon, M., Yeung, K. & Reinberg, D. Reconstitution of human TFIID activity from recombinant polypeptides: a role in TFIID-mediated transcription. *Genes Dev.* **8**, 2336-2348 (1994).
- Pugh, B. F. & Tijan, R. Mechanism of transcriptional activation by Sp1: evidence for coactivators. *Cell* **61**, 1187-1197 (1990).
- Zawel, L., Kumar, K. P. & Reinberg, D. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev.* **9**, 1479-1490 (1995).
- Tora, L. *et al.* The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell* **59**, 477-487 (1989).
- Boam, D. S., Davidson, I. & Chambon, P. A TATA-less promoter containing binding sites for ubiquitous transcription factors mediates cell type-specific regulation of the gene for transcription enhancer factor-1 (TEF-1). *J. Biol. Chem.* **270**, 487-494 (1995).
- Colgan, J. & Manley, J. L. TFIID can be rate limiting *in vivo* for TATA-containing, but not TATA-lacking, RNA polymerase II promoters. *Genes Dev.* **6**, 304-315 (1992).
- McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357-361 (1997).
- Brou, C. *et al.* Sequence-specific transactivators counteract topoisomerase II-mediated inhibition of *in vitro* transcription by RNA polymerase I and II. *Nucleic Acids Res.* **21**, 4011-4018 (1993).
- Mengus, G. *et al.* Cloning and characterization of hTAF_{II}18, hTAF_{II}20 and hTAF_{II}28: three subunits of the human transcription factor TFIID. *EMBO J.* **14**, 1520-1531 (1995).
- May, M., Mengus, G., Lavigne, A.-C., Chambon, P. & Davidson, I. Human TAF_{II}28 promotes transcriptional stimulation by activation function 2 of the retinoid X receptors. *EMBO J.* **15**, 3093-3104 (1996).
- Lavigne, A. C. *et al.* Multiple interactions between hTAF_{II}55 and other TFIID subunits. Requirements for the formation of stable ternary complexes between hTAF_{II}55 and the TATA-binding protein. *J. Biol. Chem.* **271**, 774-780 (1996).
- Besse, S., Vigneron, M., Pichard, E. & Puvion-Dutilleul, F. Synthesis and maturation of viral transcripts in herpes simplex virus type 1 infected HeLa cells: the role of interchromatin granules. *Gene Express.* **4**, 143-161 (1995).
- Wasylyk, C. & Wasylyk, B. The immunoglobulin heavy-chain B-lymphocyte enhancer efficiently stimulates transcription in non-lymphoid cells. *EMBO J.* **5**, 553-560 (1986).
- Connelly, S. & Manley, J. L. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev.* **2**, 440-452 (1988).
- Logan, J., Falck-Pedersen, E., Darnell, J. E. Jr & Shenk, T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta major-globin gene. *Proc. Natl Acad. Sci. USA* **84**, 8306-8310 (1987).
- Whitelaw, E. & Proudfoot, N. Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha-2 globin gene. *EMBO J.* **5**, 2915-2922 (1986).

Acknowledgements. We thank P. Chambon for his continuous support; A. Staub for peptide sequencing; J. M. Chipoulet for DE0.35 fractions; D. Boam, A. Bertolotti, I. Davidson, V. Dubrovskaya, X. Jacq, A. C. Lavigne, G. Mengus and M. Vigneron for reagents; D. J. Heard for critical reading of the manuscript and for discussions; I. Cheynel for expression of the TAF_{II}s and CPSF in the baculovirus system; Y. Lutz for antibodies; the Cell Culture group for HeLa cells; and R. Buchert and J.-M. Lafontaine for illustrations. J.-C.D. was supported by fellowships from the Ligue Nationale Contre le Cancer and CNRS/Région Alsace. This work was supported by grants from CNRS, INSERM, Centre Hospitalier Universitaire Régional, Ministère de la Recherche et Technologie, Fondation pour la Recherche Médicale and Association pour la Recherche contre le Cancer (L.T.) and from the NIH (J.L.M.).

Correspondence and requests for materials should be addressed to L.T. (e-mail: laszlo@titus.u-strasbg.fr).

Visualization of elongation factor Tu on the *Escherichia coli* ribosome

Holger Stark^{*†}, Marina V. Rodnina[‡], Jutta Rinke-Appel[§], Richard Brimacombe[§], Wolfgang Wintermeyer[‡] & Marin van Heel^{*}

^{*}Imperial College of Science, Medicine and Technology, Department of Biochemistry, London SW7 2AY, UK

[†]Fritz-Haber-Institut der Max-Planck-Gesellschaft, D-14195 Berlin, Germany

[‡]Institut für Molekularbiologie, Universität Witten/Herdecke, D-58448 Witten, Germany

[§]Max-Planck-Institut für Molekulare Genetik, D-14195 Berlin, Germany

The delivery of a specific amino acid to the translating ribosome is fundamental to protein synthesis. The binding of aminoacyl-transfer RNA to the ribosome is catalysed by the elongation factor Tu (EF-Tu). The elongation factor, the aminoacyl-tRNA and GTP form a stable 'ternary' complex that binds to the ribosome. We have used electron cryomicroscopy and angular reconstitution¹ to visualize directly the kirromycin-stalled ternary complex in the A site of the 70S ribosome of *Escherichia coli*. Electron cryomicroscopy had previously given detailed ribosomal structures at 25 and 23 Å (refs 2, 3) resolution, and was used to determine the position of tRNAs on the ribosome^{4,5}. In particular, the structures⁵ of pre-translocational (tRNAs in A and P sites) and post-translocational ribosomes (P and E sites occupied) were both visualized at a resolution of ~20 Å. Our three-dimensional reconstruction at 18 Å resolution shows the ternary complex spanning the inter-subunit space with the acceptor domain of the tRNA reaching into the decoding centre. Domain 1 (the G domain) of the EF-Tu is bound both to the L7/L12 stalk and to the 50S body underneath the stalk, whereas domain 2 is oriented towards the S12 region on the 30S subunit.

The ternary complex binds to the ribosome and, following codon recognition and GTP hydrolysis, delivers the aminoacyl-tRNA to the ribosomal A site. Ribosome complexes were formed on a near-natural messenger RNA by first binding fMet-tRNA^{Met} to the P site in the presence of initiation factors, then adding the antibiotic kirromycin, and finally the EF-Tu-GTP-Phe-tRNA^{Phe} complex (see Methods). The antibiotic stalls the multi-step process of A-site binding immediately after codon recognition and GTP hydrolysis, that is, EF-Tu is kept in the GTPase conformation, which is probably similar to the GTP-bound conformation, and the acceptor domain of aminoacyl-tRNA is prevented from entering the peptidyl transferase centre, presumably by remaining bound to the factor⁶⁻⁸. Electron cryomicroscopy of ice-embedded individual molecules, digitization of the micrographs, and three-dimensional image reconstruction including multivariate statistical analysis, multi-reference alignment and angular reconstitution were performed as previously described^{3,5,9}.

Figure 1a, b shows the resulting 18 Å resolution three-dimensional structure. For comparison, in Fig. 1c, d the structure of the complex obtained without kirromycin is shown at slightly lower resolution⁵; the latter complex (after GTP hydrolysis, dissociation of EF-Tu-GDP and peptide bond formation) carries tRNA^{Met} in the P site and fMet-Phe-tRNA^{Phe} in the A site ('A/P complex'). A difference map between the two structures shows significant structural differences only in the binding regions of the tRNA and the ternary complex (data not shown). In Fig. 1c, d, the A-site-bound peptidyl-tRNA, which bridges the gap between the 30S and 50S subunit, is located to the left of the L7/L12 stalk underneath the A-site finger³, leaving empty the space below. In the structure of the kirromycin-stalled

complex (Fig. 1a, b), this space contains additional density attributed to the ternary complex. The shape of the ternary complex, as revealed in this new electron microscopic reconstruction, can be directly recognized from the atomic coordinates of EF-Tu-GMPPNP-Phe-tRNA^{Phe} as derived by X-ray crystallography¹⁰ (the yeast Phe-tRNA^{Phe} in combination with the *Thermus aquaticus* EF-Tu).

The anticodon arm reaches into the 30S subunit at about the same position as in the A/P complex, whereas the acceptor arm is connected to EF-Tu; accordingly, the connection of the A-site tRNA to the 50S subunit seen in the A/P complex (Fig. 1c, d) is missing in the kirromycin-blocked complex. These features precisely reflect the biochemical properties of the latter complex, that is, an anticodon-codon contact in the decoding centre, the EF-Tu being present in the GTP-like conformation with the aminoacyl-tRNA still bound and, accordingly, no contact of the aminoacyl end of the tRNA with the 50S subunit. In the view of the ternary complex shown in Fig. 1, the body of EF-Tu lies in front of the acceptor arm of the tRNA, with the EF-Tu domain 3 reaching towards the tRNA.

In the reconstruction of the kirromycin-stalled complex, domain 1 of EF-Tu is bound to the 50S subunit in a region that, in the reconstruction, appears as a large structural element at the base of the L7/L12 stalk (Fig. 2, bottom stereo pair). This region has been implicated in factor-dependent GTP hydrolysis and probably contains proteins L11 and L10 as well as the alpha-sarcin stem loop around position 2,660 of 23S RNA, with which EF-Tu appears to interact as indicated both by functional¹¹ and by footprinting¹² data. The present reconstruction is also in line with previous data

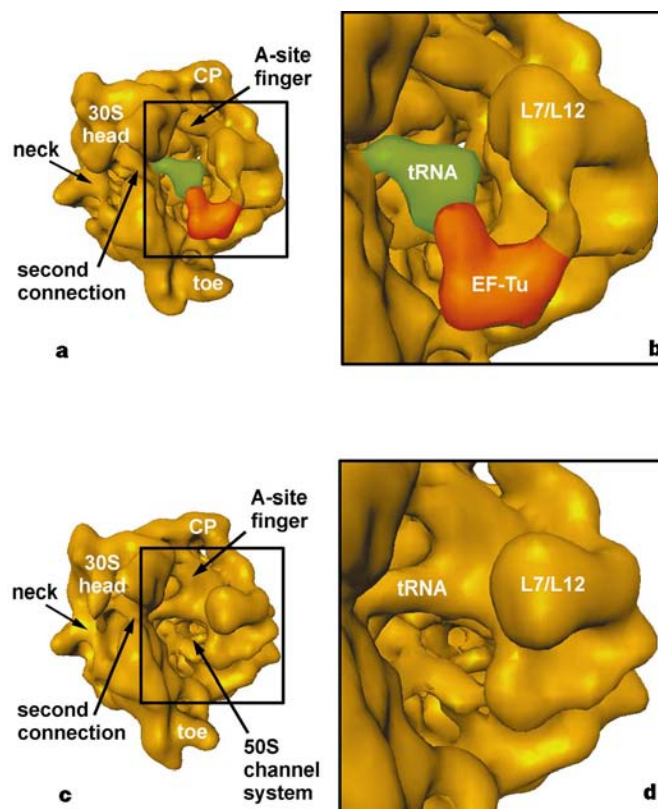


Figure 1 a, b, The 70S *E. coli* ribosome with the ternary complex locked into the A site by kirromycin. Domain 1 of the EF-Tu forms a bridge to the L7/L12 stalk on the 50S subunit and is directly adjacent to the L11/L10 site on the large subunit. Other details indicated follow earlier nomenclature⁵. c, d, The A/P state of the ribosome. After GTP hydrolysis the ribosome normally (if not stalled by kirromycin) would go into the A/P pre-translocational state carrying tRNA^{Met} in the P site and fMet-Phe-tRNA^{Phe} in the A site⁵. In the A/P state the A-site tRNA appears somewhat stretched⁵ when compared to the classic L-shape of the tRNA bound to the EF-Tu in the ternary complex.

obtained by immunoelectron microscopy¹³ that mapped EF-Tu close to the L7/L12 stalk and to proteins L10 and L11, partially overlapping the binding site of EF-G¹⁴.

Another prominent feature of the structure is the bridge between domain 1 of the EF-Tu and the L7/L12 stalk (Figs 1b and 2). This feature remains visible at a wide range of threshold levels and thus most probably represents a binding interaction. The arrangement of the ternary complex (Fig. 2, top and middle) suggests that the side of domain 1 of EF-Tu that comprises helices C and D is involved in the contact. A contact between EF-Tu and the L7/L12 stalk is in line with previous data (reviewed in ref. 15), implicating L7/L12 both in binding of protein synthesis factors and in GTPase activation, including initiation factor 2, EF-Tu, EF-G and release factor 3. Ribosomal particles that lacked L7/L12, or their C-terminal domains, or in which L7/L12 was blocked chemically or by antibodies, were impaired in factor binding and GTP hydrolysis. In keeping with these data, the present results show that EF-Tu binds to the outer end of the stalk, that is, to the C-terminal domain(s) of L7/L12 (Fig. 2).

Comparison of the L7/L12 stalk structure in the kirromycin-blocked complex and in the A/P complex (Fig. 1b, d) reveals small differences. In the EF-Tu-containing complex, the L7/L12 stalk appears bent towards EF-Tu to form the bridge between L7/L12 and the elongation factor. There is evidence suggesting that, in various functional states of the ribosome-elongation factor complex, the L7/L12 stalk assumes different conformations associated with bending in the flexible hinge region of the L7/L12 dimer¹⁶. Note that here we are studying a kirromycin-blocked complex, that is, EF-Tu is kept in the GTPase conformation immediately after GTP

hydrolysis⁸. It is conceivable, therefore, that the contact between domain 1 of EF-Tu and the C-terminal domain of L7/L12 is related to GTP hydrolysis. In an analogous reconstruction of the A-site binding complex using the non-hydrolysable GTP analogue, GMP-PNP, the L7/L12 stalk was not visible at all, presumably due to structural flexibility of the stalk (unpublished results). This finding indicates that, before GTP hydrolysis, the contact between L7/L12 and EF-Tu may be weaker or even non-existent.

On the small ribosomal subunit, immune electron microscopy suggested a contact of EF-Tu on the body of the particle on the side opposite to the platform^{13,17}. The present reconstruction of the kirromycin-stalled ternary complex shows that domain 2 of EF-Tu is oriented towards and touches (or almost touches at the current threshold level) that region of the 30S subunit (Fig. 2, top and middle). In this region of the 30S subunit, proteins S12, S5 and S4 were localized by neutron scattering¹⁸. The proteins interact with the 16S RNA in the 530 stem-loop region¹⁹, hence the placement of the latter in this part of the 30S particle¹⁹. Alternatively, crosslinking data argue in favour of placing the 530 stem-loop close to the decoding centre²⁰. In a new model, in which the 16S rRNA has been fitted to the reconstruction of Fig. 1c, the 530 stem-loop spans the gap between S12 and the decoding region²¹. In view of the importance of proteins S12, S5 and S4 as well as of the 530 stem-loop region^{22–24} for the accuracy of decoding, interaction of EF-Tu with the respective part of the 30S subunit may be of particular functional significance.

The ternary complex binding site on the 70S ribosome opens widely towards the solvent, allowing unhindered access for ternary

Figure 2 Stereo views of the ternary complex on the 70S *E. coli* ribosome seen in directions ~40° apart, all with the 30S subunit on the left and the 50S subunit on the right. **a**, Clearly discernible in this view is the bridge connecting domain 1 of the EF-Tu and the C-terminal domain of the L7/L12 stalk. Domain 2 of the EF-Tu is close to the S12, S5 and S4 proteins, and the 530 stem-loop region on the 30S subunit (see main text). **b**, The anticodon arm of the tRNA makes contact with the decoding region in the neck of the 30S subunit. Also clearly visible is the bridge from the EF-Tu to the L7/L12 stalk. **c**, In this view, the aminoacyl-tRNA is almost entirely hidden behind the L7/L12 stalk. In this direction, the contact between domain 1 of the EF-Tu and the area below the L7/L12 stalk is clearly visible.

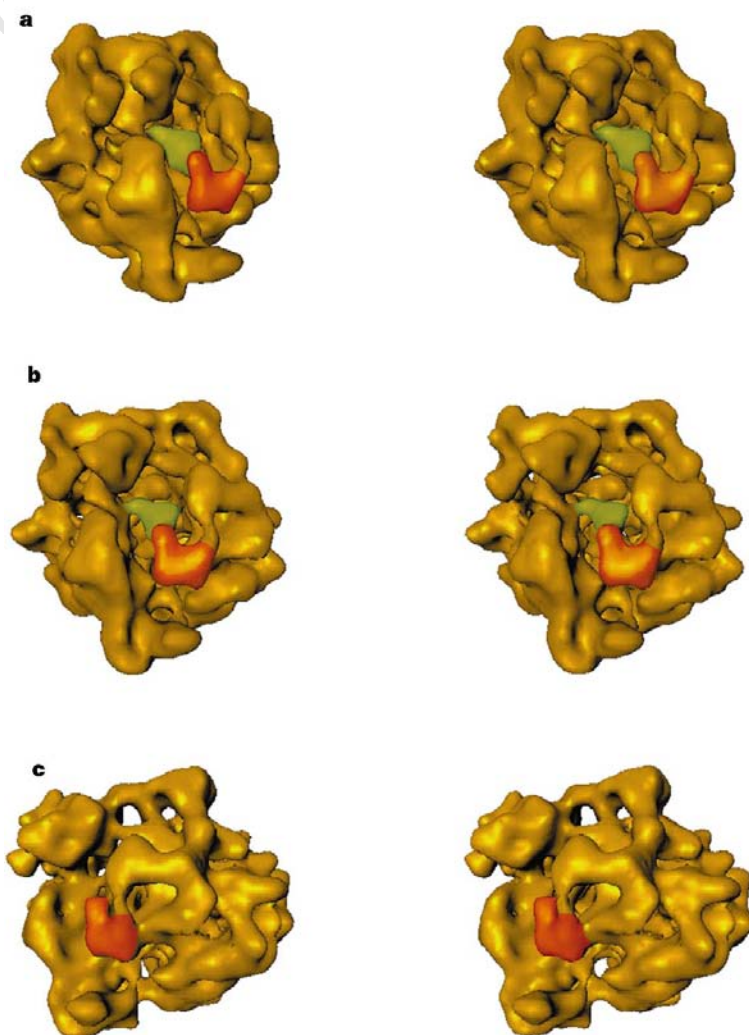
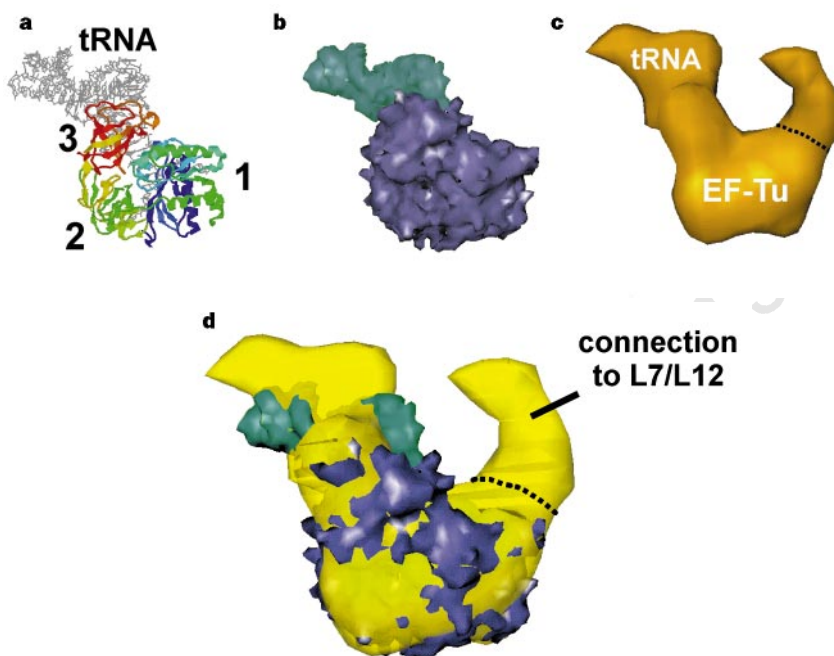


Figure 3 The crystallographic structure of the ternary complex (at 2.7 Å resolution) compared to the electron microscopical reconstruction (18 Å resolution). **a**, The ternary complex as revealed by X-ray crystallography¹⁰, with the aminoacyl-tRNA shown in grey in the background in 'stick' representation ('RASMOL'), and the EF-Tu in cartoon representation in the foreground, rainbow colour-coded along the length of the chain. The viewing direction is chosen to match that of the ternary complex within the electron microscope reconstruction shown in **c**. **b**, X-ray structure of the ternary complex in space-fill representation ('IMAGIC'). **c**, Structure of the ternary complex in the kirromycin-stalled pre-translocational 70S ribosome extracted interactively from the electron microscope 3D reconstruction. **d**, The X-ray and electron microscopical densities superimposed. The X-ray structure was aligned with respect to the electron microscopic reconstruction mainly by the position of domain 2 of the EF-Tu. Although the tRNA part of the ternary complex in the electron microscopic reconstruction is simply a low-resolution version of the X-ray structure, in the EF-Tu part of the complex some differences exist with respect to the GTP-state X-ray structure that cannot be accounted for by just the differences in resolution between the structures.



complexes to be screened in the initial selection step. Subsequent codon recognition triggers GTP hydrolysis on EF-Tu⁸. This hydrolysis leads to an extensive conformational change of the factor^{25–28} which allows the aminoacyl-tRNA to move into the A site and take part in peptide bond formation. The arrangement of the ternary complex in the kirromycin-stalled complex in comparison to the A/P complex (Figs 1 and 2) suggests that, in the absence of the antibiotic, the aminoacyl end of the tRNA is probably allowed to move freely into the A site as soon as it is released from EF-Tu-GDP, because there is empty space underneath the tRNA. The EF-Tu is located towards the solvent relative to the tRNA and will, therefore, not hinder the movement of the aminoacyl end. The movement may entail a rotation of the whole tRNA molecule, possibly in combination with a rotation of the acceptor domain relative to the anticodon domain of the tRNA.

Although there is a clear overall similarity between the X-ray structure of the ternary complex (Fig. 3a, b) and its electron microscopical *in situ* structure (Fig. 3c), some differences between the two structures (Fig. 3d) cannot be accounted for by just the differences in resolution of the techniques involved. Although the tRNA parts of the two structures are very similar, some significant differences can be seen between the EF-Tu parts of the structures and in the relative orientation of the tRNA with respect to the EF-Tu part of the ternary complex. The best overall fit between the structures was attained when the EF-Tu domain 2 areas of the reconstructions were aligned (Fig. 3d).

Apart from the slight bending of the anticodon domain relative to the body of EF-Tu, domain 3 of EF-Tu also appears shifted relative to domains 1 and 2 in the ribosome complex (Fig. 3c). It is likely that the structure of the kirromycin-blocked complex is like the transition state (GTPase state) of the complex induced by codon recognition and accommodation of EF-Tu on the ribosome^{7,8,29}; structural details may be different in different functional states. The present reconstructions, again, demonstrate the potential of the electron cryomicroscopy technique for the study of biologically functional complexes. Thus, we expect that studying further intermediate states of A-site binding will reveal the structural rearrangements of the ribosome-EF-Tu-tRNA complex during the transitions from initial binding³⁰ to the final A-site complex. □

Methods

Ribosomal initiation complexes were formed in buffer A (50 mM Tris-HCl, pH

7.5, 70 mM NH₄Cl, 30 mM KCl, 7 mM MgCl₂, 1 mM dithiothreitol) by incubating 70S ribosomes from *E. coli* MRE600 (ref. 31; 1 μM) with MFT-mRNA (120 nucleotides, coding sequence AUGUUUACG) (6 μM), *E. coli* f[3H]Met-tRNA^{Met} (3 μM; 2250 d.p.m. pmol⁻¹), initiation factors 1 (7 μM), 2 (2.3 μM) and 3 (2.3 μM), and GTP (1 mM) for 60 min at 37 °C; then kirromycin (0.1 mM) was added. The ternary complex was formed in buffer A by incubating EF-Tu (2 μM) with GTP (1 mM), phosphoenol pyruvate (3 mM) and pyruvate kinase (0.25 μg ml⁻¹) for 15 min at 37 °C, then [¹⁴C]Phe-tRNA^{Phe} was added (2 μM; 1,170 d.p.m. pmol⁻¹) and the incubation was continued for 3 min at 20 °C. The kirromycin-stalled ribosome complex was formed by mixing equal volumes each of the kirromycin-containing initiation complex and of the ternary complex. The specimen was prepared by rapidly freezing the sample on a holey carbon foil as described³. The images were taken at a defocus of 0.6 μm on a Philips CM12 electron microscope using a Gatan cryo-transfer system and cryo holder. The micrographs were digitized on the Image Science GmbH EMiL linear CCD densitometer (courtesy of Image Science) and all image processing was performed in the context of the IMAGIC image processing system⁹. The 3D reconstruction of the ternary complex ribosomes was performed using the A/P pre-translocational ribosome as a starting point for the iterative refinement procedures^{3,5,9}.

Received 11 March; accepted 16 June 1997.

- van Heel, M. Angular reconstitution: *a posteriori* assignment of projection directions for 3D reconstruction. *Ultramicroscopy* **21**, 111–124 (1987).
- Frank, J. *et al.* A model of protein synthesis based on cryo-electron microscopy of the *E. coli* ribosome. *Nature* **376**, 441–444 (1995).
- Stark, H. *et al.* The 70S *E. coli* ribosome at 23 Å resolution: fitting the ribosomal RNA. *Structure* **3**, 815–821 (1995).
- Agrawal, R. K. *et al.* Direct visualization of A-, P-, and E-site transfer RNAs in the *Escherichia coli* ribosome. *Science* **271**, 1000–1002 (1996).
- Stark, H. *et al.* Arrangement of tRNAs in pre- and posttranslational ribosomes revealed by electron cryomicroscopy. *Cell* **88**, 19–28 (1997).
- Wolf, H., Chinali, G. & Parmeggiani, A. Mechanism of the inhibition of protein synthesis by kirromycin. *Eur. J. Biochem.* **75**, 67–75 (1977).
- Mesters, J. R. *et al.* The structural and functional basis for the kirromycin resistance of mutant EF-Tu species in *Escherichia coli*. *EMBO J.* **13**, 4877–4885 (1994).
- Rodnina, M. V., Fricke, R., Kuhn, L. & Wintermeyer, W. Codon-dependent conformational change of elongation factor Tu preceding GTP hydrolysis on the ribosome. *EMBO J.* **14**, 2613–2619 (1995).
- van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
- Nissen, P. *et al.* Crystal structure of the ternary complex of Phe-tRNA^{Phe}, EF-Tu, and a GTP analog. *Science* **270**, 1464–1472 (1995).
- Tappich, W. E. & Dahlberg, A. E. A single base mutation at position 2661 in *E. coli* 23S ribosomal RNA affects the binding of ternary complex to the ribosome. *EMBO J.* **9**, 2649–2655 (1990).
- Moazed, D., Robertson, J. M. & Noller, H. F. Interaction of elongation factors EF-G and EF-Tu with a conserved loop in 23S RNA. *Nature* **334**, 362–364 (1988).
- Girshovich, A. S., Bochkareva, E. S. & Vasiliev, V. D. Localization of elongation factor Tu on the ribosome. *FEBS Lett.* **197**, 192–198 (1986).
- Girshovich, A. S., Kurtskhaliya, T. V., Ovchinnikov, Y. A. & Vasiliev, V. D. Localization of the elongation factor G on *Escherichia coli* ribosome. *FEBS Lett.* **130**, 54–59 (1981).

15. Traut, R. R. *et al.* Location and domain structure of *Escherichia coli* ribosomal protein L7/L12: Site specific cysteine crosslinking and attachment of fluorescent probes. *Biochem. Cell Biol.* **73**, 949–958 (1995).
16. Dey, D., Oleinikov, A. V. & Traut, R. R. The hinge region of *Escherichia coli* ribosomal protein L7/L12 is required for factor binding and GTP hydrolysis. *Biochimie* **77**, 925–930 (1995).
17. Langer, J. A. & Lake, J. A. Elongation factor Tu localized on the exterior surface of the small ribosomal subunit. *J. Mol. Biol.* **187**, 617–621 (1986).
18. Capel, M. S. *et al.* A complete mapping of the proteins in the small ribosomal subunit of *Escherichia coli*. *Science* **238**, 1403–1406 (1987).
19. Noller, H. F. *et al.* Structure and function of ribosomal RNA. *Biochem. Cell Biol.* **73**, 997–1009 (1995).
20. Dontsova, O. *et al.* Three widely separated positions in the 16S RNA lie in or close to the ribosomal decoding region; a site-directed cross-linking study with mRNA analogues. *EMBO J.* **11**, 3105–3116 (1992).
21. Mueller, F. & Brimacombe, R. A new model for the three-dimensional folding of *Escherichia coli* 16S ribosomal RNA. *J. Mol. Biol.* (in the press).
22. Powers, T. & Noller, H. F. Evidence for functional interaction between elongation factor Tu and 16S ribosomal RNA. *Proc. Natl Acad. Sci. USA* **90**, 1364–1368 (1993).
23. van Ryk, D. I. & Dahlberg, A. E. Structural changes in the 530 loop of *Escherichia coli* 16S rRNA in mutants with impaired translational fidelity. *Nucleic Acids Res.* **23**, 3563–3570 (1995).
24. Santer, U. V. *et al.* A mutation at the universally conserved position 529 in *Escherichia coli* 16S rRNA creates a functional but highly error prone ribosome. *RNA* **1**, 89–94 (1995).
25. Berchtold, H. *et al.* Crystal structure of active elongation factor Tu reveals major domain rearrangements. *Nature* **365**, 126–132 (1993).
26. Kjeldgaard, M., Nissen, P., Thirup, S. & Nyborg, J. The crystal structure of elongation factor EF-Tu from *Thermus aquaticus* in the GTP conformation. *Structure* **1**, 35–50 (1993).
27. Polekhina, G. *et al.* Helix unwinding in the effector region of elongation factor EF-Tu-GDP. *Structure* **4**, 1141–1151 (1996).
28. Abel, K., Yoder, M., Hilgenfeld, R. & Jurnak, F. An alpha to beta conformational switch in EF-Tu. *Structure* **4**, 1153–1159 (1996).
29. Rodnina, M. V., Fricke, R. & Wintermeyer, W. Transient conformational states of aminoacyl-tRNA during ribosome binding catalyzed by Elongation Factor Tu. *Biochemistry* **33**, 12267–12275 (1994). **Article title?**
30. Rodnina, M. V., Pape, T., Fricke, R., Kuhn, L. & Wintermeyer, W. Initial binding of the elongation factor Tu-GTP-aminoacyl-tRNA complex preceding codon recognition on the ribosome. *J. Biol. Chem.* **271**, 646–652 (1996).
31. Rodnina, M. V. & Wintermeyer, W. GTP consumption of elongation factor Tu during translation of heteropolymeric mRNAs. *Proc. Natl Acad. Sci. USA* **92**, 1945–1949 (1995).

Acknowledgements. We thank J. Nyborg for the X-ray coordinates of the ternary complex. We acknowledge support with the IMAGIC system from M. Schatz and R. Schmidt of Imag. Science GmbH. The work was supported by the Deutsche Forschungsgemeinschaft, grants to W.W., M. A. and R.B.; H.S. was financed in part by an award to R. Henderson of the MRC Laboratory for Molecular Biology, Cambridge, UK.

Correspondence and requests for materials to M.v.H. (e-mail: m.vanheel@ic.ac.uk).

Haem-ligand switching during catalysis in crystals of a nitrogen-cycle enzyme

Pamela A. Williams*†, Vilmos Fülöp*‡, Elspeth F. Garman*, Neil F. W. Saunders§, Stuart J. Ferguson*§ & Janos Hajdu*‡||

*Laboratory of Molecular Biophysics, and §Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

‡Oxford Centre for Molecular Sciences, University of Oxford, South Parks Road, Oxford OX1 3QT, UK

||Department of Biochemistry, Uppsala University, Box 576, S-75123 Uppsala, Sweden

Cytochrome *cd*₁ nitrite reductase catalyses the conversion of nitrite to nitric oxide in the nitrogen cycle^{1,2}. The crystal structure of the oxidized enzyme^{3,4} shows that the *d*₁ haem iron of the active site is ligated by His/Tyr side chains, and the *c* haem iron is ligated by a His/His ligand pair. Here we show that both haems undergo re-ligation during catalysis. Upon reduction, the tyrosine ligand of the *d*₁ haem is released to allow substrate binding. Concomitantly, a refolding of the cytochrome *c* domain takes place, resulting in an unexpected change of the *c* haem iron coordination from His 17/His 69 to Met106/His69. This step is similar to the last steps in the folding of cytochrome *c*^{5–8}. The changes must affect the redox potential of the haems, and suggest a mechanism by

which internal electron transfer is regulated. Structures of reaction intermediates show how nitric oxide is formed and expelled from the active-site iron, as well as how both haems return to their starting coordination. These results show how redox energy can be switched into conformational energy within a haem protein.

Cytochrome *cd*₁ from *Thiosphaera pantotropha* is a bifunctional enzyme that catalyses the one-electron reduction of nitrite to nitric oxide in denitrification, and the four-electron reduction of oxygen to water, which is a cytochrome oxidase reaction^{1,2,9}. Nitric oxide is a product of the main physiological reaction ($\text{NO}_2^- + 2\text{H}^+ + \text{e}^- \rightarrow \text{NO} + \text{H}_2\text{O}$), and it must be released from the *d*₁ haem despite its high affinity for iron¹⁰. The oxidation states of both types of haem iron cycle between Fe(II) and Fe(III) during nitrite reduction. The affinity of NO for a haem iron increases dramatically when Fe(III) is reduced to Fe(II) (ref. 11). The timing of reduction and the regulation and release of nitric oxide from the active site of cytochrome *cd*₁ are therefore critical if the enzyme is to avoid becoming trapped in an inactive state.

Crystals of the oxidized enzyme contain the whole dimeric molecule in the asymmetric unit³, and so the environments of the two chemically identical subunits (A and B in Fig. 1) are different in the crystal. The polypeptide chain is folded into two domains. An α -helical *c*-type cytochrome domain (residues 1–134, the *c* domain) contains the covalently attached *c* haem, the site of electron entry from donor proteins^{1,12}. The rest of the polypeptide chain (residues 135–567) forms a rigid β -propeller domain (the *d*₁-domain), containing the non-covalently bound *d*₁ haem. The axial ligands of the *d*₁ haem in the oxidized enzyme are a tyrosine from the *c* domain (Tyr 25) and a histidine from the *d*₁ domain (His 200). The *c* haem has two histidine ligands (His 17 and His 69) and so is different from most other cytochromes *c*, which have His–Met iron ligation.

Time-resolved structural studies using single-crystal microspectrophotometry^{13,14} and X-ray crystallography show the enzyme undergoing major structural rearrangements during catalysis. Reaction intermediates during nitrite reduction were trapped by freeze-quenching crystals at 90K, and monochromatic data sets were collected to high resolution from these intermediates.

The *d*₁ haem has to be reduced for catalysis to occur. It receives electrons from the *c* haem, which in turn can receive electrons from external electron shuttle proteins such as azurin, pseudoazurin or *c*-type cytochromes (see ref. 1). The electron transfer between the soluble redox partners and the *c* haem of the enzyme is diffusion limited. The electron is then transferred internally from the *c* haem to the *d*₁ haem, the site of catalytic activity; this transfer is slow (0.1–0.5 s^{–1})^{15,16}, and seems to be the rate-limiting step in the catalytic reaction.

Crystalline cytochrome *cd*₁ can be reduced by small-molecule reducing agents which, unlike physiological electron donors, can diffuse into the lattice. Crystals of oxidized cytochrome *cd*₁ (ref. 17) were reduced using either sodium dithionite or electrochemically reduced methyl viologen as reductant. The brownish-red crystals of the oxidized enzyme turn green upon reduction (see Fig. 2a for time course).

Monochromatic X-ray diffraction data were collected at room temperature from a reduced crystal and the structure was refined (Table 1). Reduction caused a moderate decrease in the maximum resolution relative to oxidized crystals. Comparison of the oxidized and reduced structures of cytochrome *cd*₁ (Fig. 1a, b) shows that the postulated displacement³ of Tyr 25 from the *d*₁ haem has occurred on reduction (Fig. 3), together with an unexpected reorganization of the *c* domain (Fig. 4). His 17 has been replaced as a ligand to the *c* haem iron by Met 106, from loop 99–116 between helices 5 and 6 (for notations see ref. 4). In the oxidized form, the first residue of the *c* domain with unambiguous electron density is Asp 9 in both subunits, but in the reduced form the first residue is Asp 36 in subunit A and Glu 26 in subunit B, leaving His 17 and Tyr 25

† Present address: Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

Figure 1 Architecture of oxidized (**a**) and reduced (**b**) cytochrome *cd*₁ nitrite reductase from *Thiosphaera pantotropha*. The two chemically identical subunits (A and B) are in different environments. The protein is rainbow coloured, blue at the N terminus. Residues 1–134 form a c-type cytochrome domain (mainly blue) containing the c haem, the site of electron entry. Residues 135–567 form a rigid propeller domain, containing the non-covalently bound *d*₁ haem, the site of nitrite reduction. In the oxidized enzyme (**a**), the first 8 residues are disordered in both subunits. In the reduced enzyme (**b**), the first 35 residues of subunit A and the first 25 residues of B are disordered. Residues 63–121 of subunit A are in weak density, and were modelled at half occupancy. The axial ligands of the *d*₁ haem in (**a**) are Tyr 25 from the c domain, and His 200 from the *d*₁ domain. The c haem has His 17–His 69 ligation. After reduction, Tyr 25 leaves the *d*₁ haem, and the c haem changes ligation from His 17–His 69 (**a**) to Met 106–His 69 (**b**).

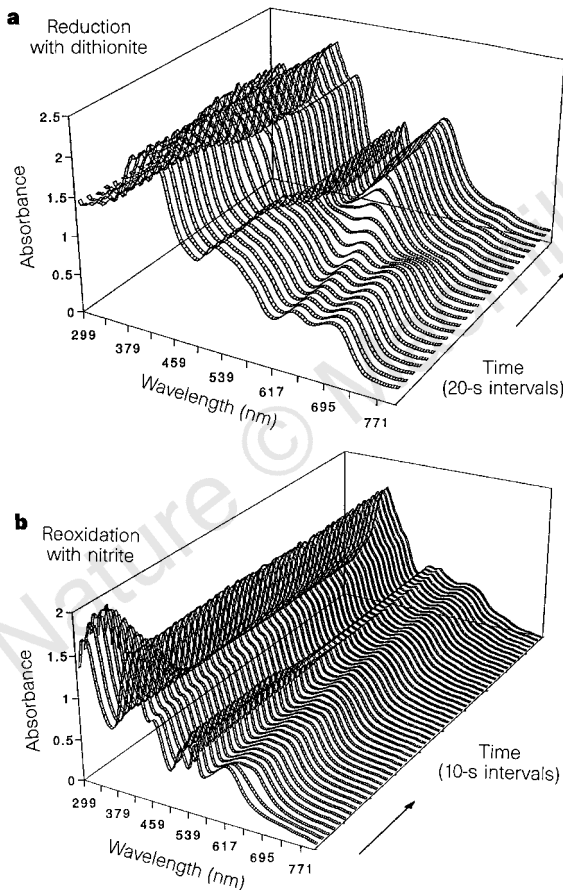
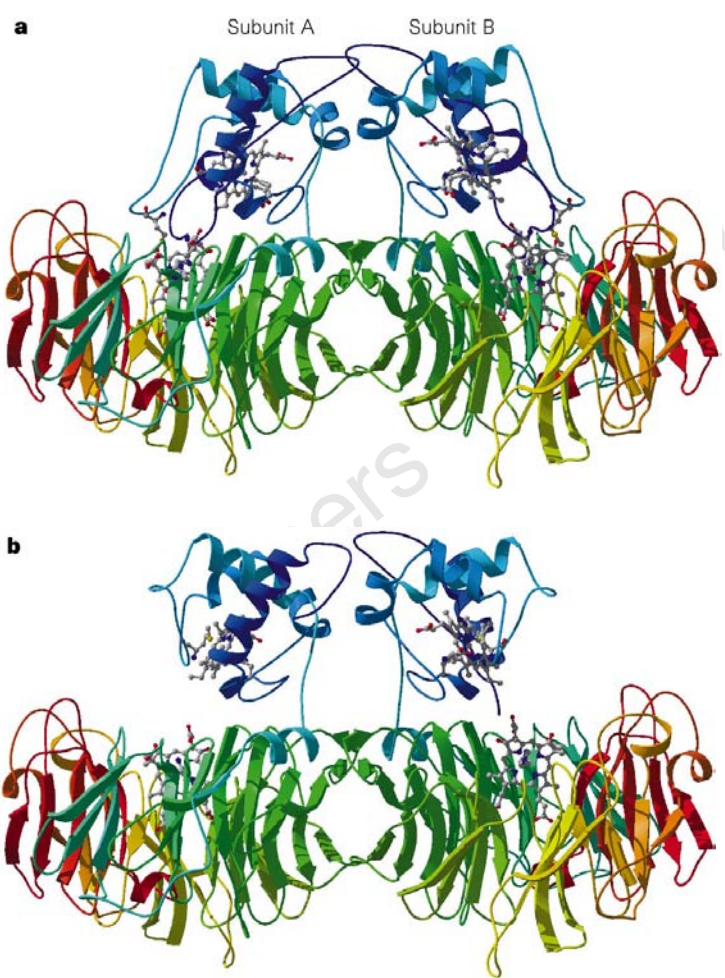


Figure 2 Spectral changes in single crystals of cytochrome *cd*₁ during reduction by dithionite (**a**) and reoxidation by nitrite (**b**). The spectrum of the fully oxidized enzyme is at the front in **a**, and that of the fully reduced enzyme is at the front in **b**. Spectra were recorded with a portable microspectrophotometer (4DX Systems, Uppsala). The c and *d*₁ haems have distinct spectral signals (oxidized c haem, 411 and 520 nm; reduced c haem, 417, 520 and 548 nm; oxidized *d*₁ haem, 470 and 640 nm; reduced *d*₁ haem, 456 and 650 nm (ref. 9)).

Table 1 Data collection and refinement statistics

Compound	Temperature (K)	Resolution (Å)	No. of observations	Unique reflections	Completeness (%)	R_{merge}^* (%)	$R_{\text{cryst}}^\dagger$ (%)	$R_{\text{free}}^\ddagger$ (%)	No. of non-H atoms
Reduced enzyme (dithionite)	293	20–2.2	461,872	80,774	94.0	6.2	16.3	20.1	11,969
Reduced enzyme (methyl viologen)	90	20–1.6	622,135	147,330	94.1	3.2	19.0	21.9	13,673
Crystal 1 soaked in nitrite (Fig. 5a, b)	90	20–1.8	299,460	98,949	90.2	4.6	18.1	20.0	12,827
Crystal 2 soaked in nitrite (Fig. 5c, d)	90	20–1.8	303,438	101,329	94.4	2.9	17.6	19.7	14,013

* $R_{\text{merge}} = \frac{\sum_i \sum_j |I_{ij} - \langle I_i \rangle|}{\sum_i \sum_j I_{ij}} \times 100$.
† $R_{\text{cryst}} = \frac{\sum_i |F_{\text{obs}} - F_{\text{calc}}|}{\sum_i F_{\text{obs}}} \times 100$.
‡ Defined in ref. 30.

disordered in the reduced crystals. When dithionite was the reducing agent, sulphur dioxide, a by-product, was bound to the d_1 haem (Fig. 3a). With methyl viologen as reductant, no by-product was visible on the haem iron (Fig. 3b), but otherwise the same structure was observed. The conformational changes in the protein and the change in protein haem ligands on reduction are thus independent of the reducing agent used.

In contrast to the c domain, the general architecture of the eight-bladed propeller structure of the d_1 domain was unaltered on reduction. The root-mean-squared (r.m.s.) deviation between $C\alpha$

positions of the d_1 domain in the oxidized and reduced models is 0.38 and 0.35 Å for subunits A and B, respectively. Reduction triggers a change in the geometry of the d_1 haem, which becomes distorted, adopting a saddle shape. The iron–iron distance between the c and d_1 haems increases slightly on reduction (from 20.6 to 21.2 Å), and the d_1 haem becomes exposed to bulk solvent (and nitrite). The increased water content of the intervening medium between the two domains after reduction is likely to affect electron transfer rates between the c and d_1 haems. In the oxidized form there are 18 hydrogen bonds between the c and d_1 domains in subunit B

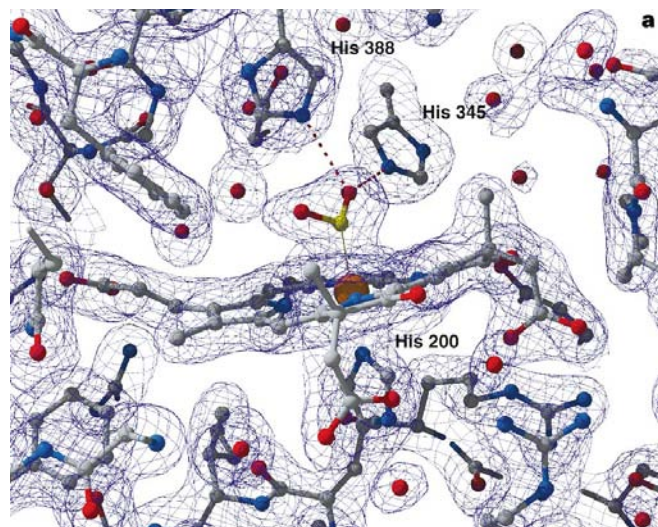
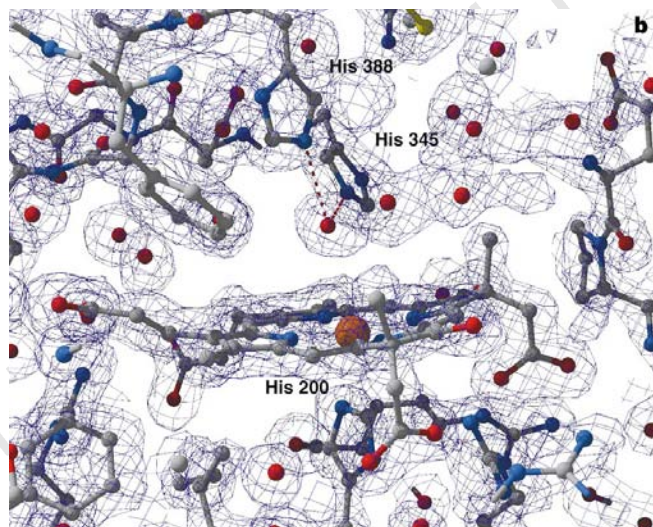


Figure 3 Electron density maps around the active site d_1 haem in reduced cytochrome cd_1 . **a**, Reduction was by dithionite, and the resulting bound SO_2 molecule is apparent. **b**, Reduction was by methyl viologen when no bound ligand



is visible directly above the haem iron; a water molecule hydrogen bonded to His 345 and His 388 can be fitted to a density above the NA nitrogen atom of the haem ring. Broken lines, selected hydrogen bonds.

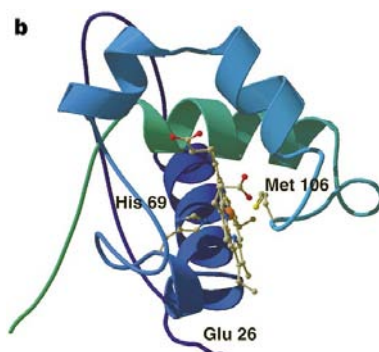
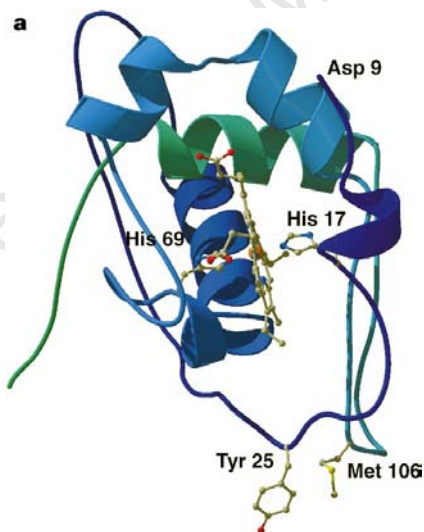
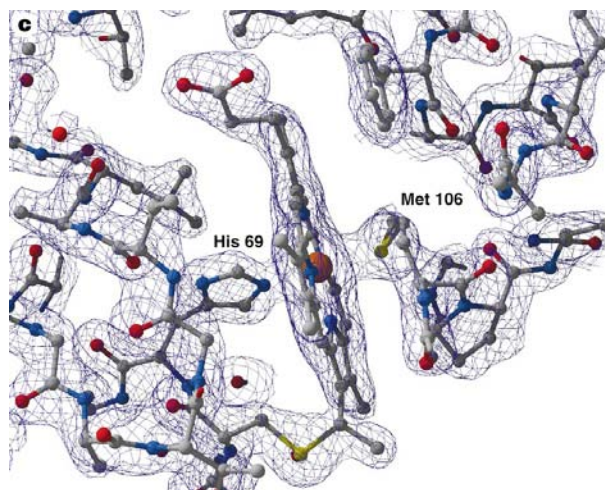


Figure 4 Structures of the c -type cytochrome domain in subunit B of oxidized (**a**) and reduced (**b**) cytochrome cd_1 . The first 8 residues are disordered in **a**, and the first 25 in **b**. The colours are as Fig. 1. **c**, Electron-density map around the reduced c haem showing His–Met ligation, contoured at 1σ , where σ represents the r.m.s. electron density for the unit cell.



and 16 in subunit A⁴, and on reduction all 16 in subunit A and 6 in subunit B are broken. Only one residue (Asn 108)⁴ in the loop containing Met 106 participates in the hydrogen-bonding network between the two domains that breaks on reduction. Met 106 makes hydrophobic interactions at the domain interface in the oxidized enzyme. The breakage of more hydrogen bonds in the reduced subunit A reflects its different environment from subunit B in the monoclinic crystal. Previous studies¹⁸ indicated a large conformational rearrangement following reduction. Despite the scale of this conformational change (Figs 1 and 4), the crystal does not crack because of the rigidity of the *d*₁ domains, their conformational independence of redox state, and their dominant contribution to lattice contacts.

These rearrangements may result from the different binding propensities of sulphur, nitrogen and oxygen ligands to oxidized and reduced haem iron^{19,20}, and the free-energy change associated with the reduction of the *c* and *d*₁ haem centres. The reduction of the *c* haem may result in a large enough decrease in affinity for His 17 to dissociate and be replaced by Met 106, the sulphur atom of which stabilizes the Fe(II) state of haem relative to the nitrogen of His 17. This could be achieved by movement of the loop that extends from residues 99–116 and connects helices 5 and 6, which themselves do not move much. Such movement of the loop requires the breaking of seven internal hydrogen bonds within the *c* domain of the oxidized enzyme and one inter-domain hydrogen bond between Trp 522 of the *d* domain and Asn 108 in the loop containing Met 106

in the *c* domain. These breaks are accompanied by the formation of eight new hydrogen bonds within the *c* domain of the reduced enzyme, and the loop region losing its hydrophobic contact with the *d*₁ haem domain. Dissociation of His 17 from the *c* haem would then trigger dissociation of Tyr 25 from the *d*₁ haem and the breakage of many hydrogen bonds between the two domains. Alternatively, the reduction of the *d*₁ haem could trigger the dissociation of Tyr 25, leading to a weaker binding of His 17 to the *c* haem, so that Met 106 would replace His 17.

Some of the spectral changes occurring during nitrite reduction by cytochrome *cd*₁ in the crystalline state are shown in Fig. 2b. Several spectrally different intermediates are formed with life times of 20–5,000 s at 18 °C. The reoxidized state of the enzyme could be regained in these experiments, and a single crystal could undergo several cycles of reoxidation by nitrite. This establishes that the crystals are catalytically active, and do not form a long-lived dead-end complex after exposure to nitrite.

The Laue method could not be used to collect interpretable data from reaction intermediates owing to the increased mosaicity of crystals in the act of catalysis. Our results are from high-resolution monochromatic studies on crystals freeze-quenched during nitrite reduction. The structures of the reduced and oxidized forms of the enzyme (the initial and final states in experiments to follow catalysis) remained unaltered on cooling to 90 K.

After reduction with dithionite, crystals of cytochrome *cd*₁ were soaked in nitrite and freeze-quenched at 90 K. Spectra recorded

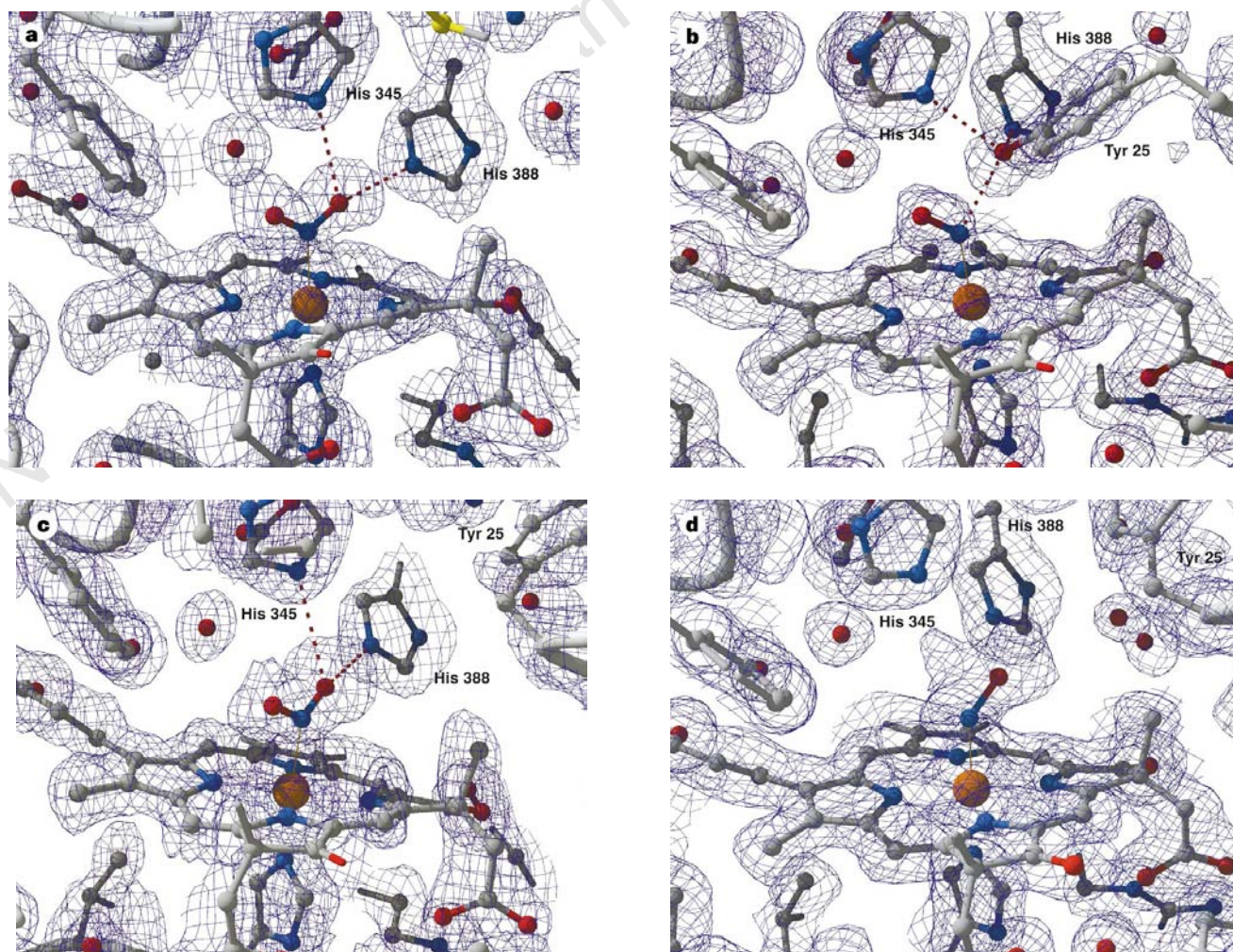


Figure 5 Structures of the *d*₁ haem with substrate (nitrite) and product (nitric oxide) bound in two different crystals of cytochrome *cd*₁. **a, b**, Crystal 1 was frozen after 1 min in nitrite; nitrite bound to subunit A (**a**) and nitric oxide bound to subunit

B (**b**). **c, d**, Crystal 2 was frozen after 3 min in nitrite with bound ligands on the haem iron of subunits A (**c**) and B (**d**). The electron density in **d** does not allow unambiguous identification of bound ligand(s).

before and after X-ray data collection at 90 K showed no change in the properties of the frozen crystals during data collection. Electron-density maps from two crystals provided structures of the enzyme complexed to either nitrite or nitric oxide (Fig. 5). The bound ligands were identified on omit maps from high-resolution data. Nitrite was seen bound by its nitrogen to the d_1 haem in subunit A in two different crystals (Fig. 5a, c). The two oxygen atoms of nitrite point approximately towards the CHB and CHD carbons of the d_1 haem ring (for notations see ref. 4). Tyr 25 is absent from the structure in one intermediate (Fig. 5a), but some of the density for Tyr 25 can be seen in the second, away from the haem plane near Thr 386 (Fig. 5c). In both of these nitrite-bound structures, one of the oxygens of nitrite is positioned suitably for hydrogen bonding to the $N^{\epsilon 2}$ atoms of His 345 and His 388 (oxygen to $N^{\epsilon 2}$ distances are 2.9–3.2 Å; Fig. 5a, c). This finding supports the previous proposal³ that these two histidines provide protons to the nitrite oxygen atom destined to become the product water molecule.

In one crystal, subunit B had nitric oxide bound to the d_1 haem (Fig. 5b), but the other reaction product, water, was not present. Nitric oxide is bound in a bent conformation to the d_1 haem iron, with an Fe–NO bond angle of 131° (Fig. 5b). This is similar to those observed for nitric oxide bound to Fe(II) haemoglobin²¹ and Fe(III) cytochrome *c* peroxidase²². The Fe–N distance is 2.0 Å (Fig. 5b, d), longer than the 1.7 Å reported for the NO–Fe(II)–haemoglobin complex²¹ from which NO dissociates extremely slowly. The simplest interpretation of Fig. 5b is that it represents the Fe(III)–NO complex. The distance between the iron of the d_1 haem and the $N^{\epsilon 2}$ atom of its His 200 ligand is not increased much when NO is bound¹, in contrast to the similar bond in the very strong Fe(II)–haemoglobin–NO complex, and so it supports the contention that Fig. 5b represents the Fe(III)–NO complex from which NO dissociation would be relatively easy. This is also supported by the shape of the d_1 haem in the NO complexes, which has the form adopted in the oxidized enzyme.

Subunit B of the second crystal clearly had a liganded d_1 haem (Fig. 5d). The density for the ligand could be modelled in more than one way, and probably represents a mixture of various species. Tyr 25 is well ordered, but points away from the haem iron. However, the shape of the haem ring and the iron–His 200 distance are similar to those in Fig. 5b, suggesting that the bound ligand is NO in multiple conformations.

Three distinct positions of Tyr 25 in different forms of the enzyme are shown in Fig. 6. In each position, the phenolate oxygen participates in a distinct pattern of hydrogen bonding. In the unliganded reduced state of the enzyme (not shown), Tyr 25 is completely disordered. Fragmented density for Tyr 25 is visible near

Thr 386 in the structure corresponding to Fig. 5c in which nitrite is bound. Well-defined density for this tyrosine with hydrogen bonding to Thr 386 is visible in the structure corresponding to Fig. 5d and shown in yellow in Fig. 6. When nitric oxide was bound to the d_1 haem (Fig. 5b (red, Fig. 6)) the phenolate oxygen of Tyr 25 occupied a position adjacent to a water molecule in the oxidized enzyme (blue, Fig. 6). The tyrosine in this position has hydrogen bonds to His 345 and His 388, and seems poised to return to ligation with the iron of the d_1 haem, as in the oxidized enzyme (blue). We propose that we have captured Tyr 25 at different stages in its return to the coordination sphere of the d_1 haem iron, in a process that may help to displace bound nitric oxide. Distinct sets of interactions define the conformational states for the tyrosine. The return of Tyr 25 may therefore involve distinct conformers rather than a smooth/uncontrolled pathway.

The structures of nitrite or nitric oxide bound to the d_1 haem demonstrate that this is the catalytic centre of cytochrome *cd*₁ nitrite reductase. They also indicate that the substrate and product bind to the iron of the d_1 haem by nitrogen rather than oxygen atoms. In contrast, binding of nitrite by oxygen to a copper centre of another type of nitrite reductase has been proposed²³. To interpret our observations, and the corresponding liganded states of the *c*-type cytochrome centre, we must consider the number of reducing equivalents held by the enzyme when the reductant was replaced by nitrite. When nitrite was introduced, the enzyme was in the fully reduced form, with each subunit carrying two reducing equivalents, one on the *c*-type haem and one on the d_1 -type haem. Nitrite must have bound to this reduced form of the enzyme. Subsequent electron transfer to the nitrite and loss of water would generate an enzyme–product species in which the d_1 haem should be oxidized and the *c*-type haem reduced. Allowing the d_1 haem to be re-reduced while NO is bound would be catalytically disadvantageous³ because NO has a much higher affinity for reduced than for oxidized haem^{10,11}. The availability of the second reducing equivalent on each subunit makes possible a second cycle of nitrite binding and nitric-oxide formation within the crystal. Thus each structure in Fig. 5 could be a form of the enzyme undergoing either the first or second round of reaction in any of the subunits.

Three of the intermediate structures (Fig. 5b–d) showed that the haem iron coordination in the *c* haem domain had returned to the His/His coordination characteristic of the oxidized form of the enzyme. In the fourth case (Fig. 5a), the electron density for the *c* haem domain of subunit A was too poorly resolved for assignment of the haem ligands. The assumption that the *c* haem is in the Fe(III) state in the resolved cytochrome *c* domains implies that Fig. 5c shows the second nitrite ion to bind. In Fig. 5b the nitric oxide may

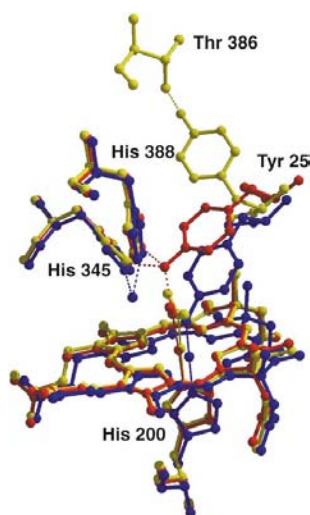


Figure 6 Positions of Tyr 25 above the d_1 haem during catalysis. Tyr 25 is not visible in electron-density maps of the fully reduced enzyme. Yellow, position of Tyr 25 in Fig. 5d in which a mixture of nitrite and nitric oxide can be fitted as ligands. Red, as Fig. 5b with nitric oxide bound, and the phenolate oxygen occupying a position of a water molecule that interacts with Tyr 25, His 345 and His 388 in the oxidized enzyme. Blue, position of Tyr 25 in the oxidized enzyme when it is bound to the haem iron³.

be bound to either a reduced or an oxidized d_1 haem. In the latter case the nitric oxide would be the second molecule to be formed. In the former case, a putative dead-end complex between NO and Fe(II)- d_1 -haem would be present. However, spectrophotometry showed that such crystals returned to the fully oxidized state, implying that the structure in Fig. 5b is NO-Fe(III)- d_1 -haem, in agreement with the observed bond lengths from iron to NO and His200, and the fact that the d_1 haem has a conformation corresponding to the Fe(III) form. The proximal position of Tyr 25 in Fig. 5b suggests that it will assist the displacement of nitric oxide from this state of the d_1 haem as postulated³. Alternatively, Tyr 25 may assist the destruction of a dead-end NO-Fe(II)- d_1 -haem complex, should it be formed by accident.

Support for the suggestion³ that Tyr 25 returns to the iron of the d_1 haem in a catalytic step designed to displace bound nitric oxide comes from the observation of Tyr 25 near the d_1 haem when nitric oxide is bound (Fig. 5b, and red structure in Fig. 6). Such a role for Tyr 25 naturally implies that coordination of the c haem iron changes between His-His and His-Met in each catalytic cycle. In general, His-His coordinated c haem has a lower redox potential than His-Met coordination^{19,20}, so the driving force for electron transfer from the His-Met centre would be lower than from the His-His centre. Correspondingly, a His-His coordinated c haem iron may be harder to reduce, enabling the enzyme to delay electron transfer to the d_1 haem until nitric oxide has departed. Redox state-dependent changes in the water content of the intervening medium (Fig. 1) are expected to enhance this effect.

If the His/His ligation is a resting state of the enzyme^{1,24}, then the rearrangement in the c haem domain on reduction would reflect a conformational change-dependent activation. However, the reversion to the His-His ligation in the crystal structures with ligands bound indicates that His/His is involved in catalytically competent forms of the enzyme.

The structural change between oxidized and reduced cytochrome cd_1 shows how redox energy can be switched into conformational energy within a haem protein, a phenomenon of general importance²⁵⁻²⁷. Similar switching occurs in the refolding of mitochondrial cytochrome c , which involves a His-His coordinated haem c on the way to the final His-Met coordinated structure⁵⁻⁸. The displacement of His by Met in cytochrome c is rate limiting in the last stages of the folding process near neutral pH. A similar change of ligation takes place in cytochrome cd_1 at the c haem during catalysis. Regions containing the removable haem ligands (His 17 and Met 106) are flexible and enable the side chains of His 17 and Met 106 to explore the vicinity of the c haem iron in the enzyme. The results suggest that elements in the folding process of a protein may be used in structural transitions required for functioning of the protein. □

Methods

Crystallization. Cytochrome cd_1 from *T. pantotropha* was purified⁹ and crystallised¹⁷ as described previously. The crystals belong to space group $P2_1$ with unit cell dimensions $a = 106.4 \text{ \AA}$, $b = 60.6 \text{ \AA}$, $c = 100.2 \text{ \AA}$, $\beta = 112.3^\circ$ (at 90K). The crystallographic asymmetric unit contains a dimer, and has a solvent content of 48% by volume.

Flow cell techniques. Crystals of $\sim 0.8 \times 0.3 \times 0.2 \text{ mm}^3$ were wedged in a tapered quartz capillary (0.5–0.2 mm diameter) filled with a synthetic mother liquor, containing 2.5 M ammonium sulphate, 0.1 M potassium phosphate/KOH, pH 7.0. Reduction of cytochrome cd_1 in the crystal was initiated by pumping 16 mM dithionite or 20 mM electrochemically reduced methyl viologen in this solution past the crystal at room temperature with a motor-driven syringe pump. The flow rate was 0.3 ml h^{-1} , ensuring a complete change of solution around the crystal every second. Oxidation of reduced cytochrome cd_1 was triggered by pumping an oxygen-free solution of 20 mM potassium nitrite in the synthetic mother liquor past the crystal.

Single-crystal microspectrophotometry. Spectra were recorded with a portable microspectrophotometer¹³ using an X-ray camera or an optical

goniometer. For comparative kinetic studies, the crystal was kept in the same orientation for each spectrum recorded. Spectra were taken in 10- or 20-s intervals from an area of 0.05 mm diameter with recording times of 0.16 s per spectrum. A reference spectrum was recorded from around the crystal before measurements.

Quench-freezing. The cryotechnique requires an open system; a higher concentration of dithionite (40 mM) and larger volumes of reagents (1 ml) were used to reduce crystals. All solutions contained 15% glycerol as a cryoprotectant, and were purged with argon before use. When a crystal appeared green from all directions, it was transferred to a synthetic mother liquor (15% glycerol, 2.5 M ammonium sulphate, 0.1 M potassium phosphate, pH 7.0), containing 20 mM potassium nitrite as substrate. After 1–3 min in the nitrite solution at 18°C , crystals were frozen at 90 K in a stream of cold N_2 gas, and spectra were recorded. The frozen crystals could be stored in liquid nitrogen for many months without measurable change in their spectra.

X-ray data collection and structure refinement. Data to 1.5–1.8 Å resolution were collected from frozen crystals at 90 K with monochromatic X-rays (wavelength 0.88 Å) on beam line 3 of ESRF (Grenoble), using a 30-cm MAR image plate detector. Room temperature data to 2.0 Å resolution on reduced crystals were collected on 18-cm MAR image plate detectors either at home (Cu Kα radiation) or at station 9.5 of SRS Daresbury (wavelength 0.877 Å). Data were processed and structures refined using standard procedures (protocols available from the authors on request). The refinement was based on the room-temperature 1.55 Å resolution model¹. The model was corrected at four places to reflect the gene sequence-derived protein sequence⁴, and the two independent molecules were refined as rigid-body groups against the low-temperature data. The model was subjected to simulated annealing refinement using weak non-crystallographic symmetry restraints, gradually extending the resolution from 2.2 Å to the final resolution of the data sets in three cycles. A bulk solvent correction allowed all measured data to be used. In the final stages the non-crystallographic symmetry restraint was released and water molecules were added to the atomic model at the positions of large positive peaks ($>3\sigma$) in the difference electron density, at places where the resulting water molecule fell into an appropriate hydrogen-bonding environment. Restrained isotropic temperature factor refinements were carried out for each individual atom. Figures were drawn by Molscript²⁸ modified by R. Esnouf and rendered with Raster3D²⁹.

Received 4 April; accepted 13 June 1997.

- Berks, B. C., Ferguson, S. J., Moir, J. W. B. & Richardson, D. J. Enzymes and associated electron transport systems that catalyse the respiratory reduction of nitrogen oxides and oxyanions. *Biochim. Biophys. Acta* **1232**, 97–173 (1995).
- Averill, B. A. Dissimilatory nitrite and nitric oxide reductases. *Chem. Rev.* **96**, 2951–2964 (1996).
- Fülöp, V., Moir, J. W. B., Ferguson, S. J. & Hajdu, J. The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome cd_1 . *Cell* **81**, 369–377 (1995).
- Baker, S. C. *et al.* Cytochrome cd_1 structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to β -propeller folds. *J. Mol. Biol.* **269**, 440–455 (1997).
- Elove, G. A., Bhuyan, A. K. & Roder, H. Kinetic mechanism of cytochrome- c folding involvement of the heme and its ligands. *Biochemistry* **33**, 6925–6935 (1994).
- Colon, W., Elove, G. A., Wakem, L. P., Sherman, F. & Roder, H. Side-chain packing of the N-terminal and C-terminal helices plays a critical role in the kinetics of cytochrome- c folding. *Biochemistry* **35**, 5538–5549 (1996).
- Takahashi, S. *et al.* Folding of cytochrome c initiated by submillisecond mixing. *Nature Struct. Biol.* **4**, 44–50 (1997).
- Yeh, S.-R., Takahashi, S., Fan, B. & Rousseau, D. L. Ligand exchange during cytochrome c folding. *Nature Struct. Biol.* **4**, 51–56 (1997).
- Moir, J. W. B., Baratta, D., Richardson, D. J. & Ferguson, S. J. The purification of a cd_1 -type nitrite reductase from, and the absence of a copper-type nitrite reductase form, the aerobic denitrifier *Thiosphaera pantotropha*—the role of pseudozurin as an electron-donor. *Eur. J. Biochem.* **212**, 377–385 (1993).
- Sharma, V. S., Taylor, T. G., Gardiner, R. & Mizukami, H. Reaction of nitric oxide with heme proteins and model compounds of hemoglobin. *Biochemistry* **26**, 3837–3842 (1987).
- Ribeiro, J. M. C., Hazzard, J. H. M., Nussenzweig, R. H., Champagne, D. E. & Walker, F. A. Reversible binding of nitric oxide by a salivary heme protein from a bloodsucking insect. *Science* **260**, 539–541 (1993).
- Williams, P. A. *et al.* Pseudospecific docking surfaces on electron transfer proteins as illustrated by pseudozurin, cytochrome $c550$, and cytochrome cd_1 nitrite reductase. *Nature Struct. Biol.* **2**, 975–982 (1995).
- Hadfield, A. T. & Hajdu, J. A fast and portable micro-spectrophotometer for time-resolved X-ray diffraction experiments. *J. Appl. Cryst.* **26**, 839–842 (1993).
- Mozzarelli, A. & Rossi, G. L. Protein function in the crystal. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 343–365 (1996).
- Silverstrini, M. C., Tordi, M. G., Antonini, E. & Brunori, M. The kinetics of electron transfer between *Pseudomonas aeruginosa* cytochrome $c551$ and its oxidase. *Biochem. J.* **203**, 445–451 (1982).
- Silverstrini, M. C., Tordi, M. G., Musci, G. & Brunori, M. The reaction of *Pseudomonas* nitrite reductase and nitrite. *J. Biol. Chem.* **265**, 11783–11787 (1990).

17. Fülöp, V., Moir, J. W. B., Ferguson, S. J. & Hajdu, J. Crystallisation and preliminary crystallographic study of cytochrome *cd*, nitrite reductase from *Thiosphaera pantotropha*. *J. Mol. Biol.* **232**, 1211–1212 (1993).
18. Berger, H. & Wharton, D. C. Small angle X-ray scattering studies of oxidised and reduced cytochrome oxidase from *Pseudomonas aeruginosa*. *Biochim. Biophys. Acta* **622**, 355–359 (1980).
19. Moore, G. R. & Pettigrew, G. W. *Cytochromes c: Evolutionary, Structural and Physicochemical Aspects* (Springer, Berlin, 1990).
20. Pettigrew, G. W. & Moore, G. R. *Cytochromes c: Biological Aspects* (Springer, Berlin, 1987).
21. Harutunyan, E. H. *et al.* The binding of carbon monoxide and nitric oxide to leghaemoglobin in comparison with other haemoglobins. *J. Mol. Biol.* **264**, 152–161 (1996).
22. Edwards, S. L., Kraut, J. & Poulos, T. L. Crystal structure of nitric oxide inhibited cytochrome-c peroxidase. *Biochemistry* **27**, 8074–8081 (1988).
23. Adman, E. T., Godden, J. W. & Turley, S. The structure of copper nitrite reductase from *Achromobacter cycloclastes* at five pH values, with NO₂ bound and with type II copper depleted. *J. Biol. Chem.* **270**, 27458–27474 (1995).
24. Williams, P. A. thesis, Oxford Univ. (1996).
25. Poulos, T. L. Ligands and electrons and haem proteins. *Nature Struct. Biol.* **3**, 401–403 (1996).
26. Wittung, P. & Malmstrom, B. G. Redox-linked conformational changes in cytochrome c oxidase. *FEBS Lett.* **388**, 47–49 (1996).
27. Pascher, T., Chesick, J. P., Winkler, J. R. & Gray, H. B. Protein folding triggered by electron transfer. *Science* **271**, 1558–1560 (1996).
28. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
29. Merritt, E. A. & Murphy, M. E. P. Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
30. Brünger, A. T. The free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–474 (1992).

Acknowledgements. We thank the ESRF and SRS Daresbury for data collection facilities; the EMBL outstation, Grenoble, for use of an image plate detector; M.L.D. Page for expert advice; R. Bryan and R. Esnouf for computing; K. Harlos for help with in-house data collection; F. Armstrong and J. Hirst for providing electrochemically reduced methyl viologen. This work was supported by MRC, BBSRC and EU-BIOTECH. The Oxford Centre for Molecular Sciences is funded jointly by BBSRC, EPSRC and MRC. N.E.W.S. was supported by a Wellcome Trust prize studentship. V.F. is a Royal Society university research fellow.

Correspondence and requests for materials should be addressed to P.A.W. (e-mail: pamela@scripps.edu), V.F. (e-mail: vilmos@biop.ox.ac.uk) or J.H. (e-mail: janos@xray.bmc.uu.se).

errata

The yeast genome directory

Nature **387** (suppl.) (1997)

In the list of authors given on page 5 of this supplement, the names of some authors were omitted or misspelled (asterisks). These were: R. Altman*; W. Arnold*; M. de Haan*; K. Hamberg; K. Hinni; L. Jones; W. Kramer; H. Küster*; K. C. T. Maurer*; D. Niblett; N. Paricio*; A. G. Parle-McDermott*; C. Rebischung; C. Richards; L. Rifkin*; J. Robben; C. Rodrigues-Pousada*; I. Schaaff-Gerstenschläger*; P. H. M. Smits*; Y. Su*; Q. J. M. van der Aart*; J. C. van Vliet-Reedijk*; A. Wach; M. Yamazaki*. □

Measurements of elastic anisotropy due to solidification texturing and the implications for the Earth's inner core

Michael I. Bergman

Nature **389**, 60–63 (1997)

Owing to a typographical error, this Letter appeared under the title “Measurements of electric anisotropy due to solidification texturing and the implications for the Earth's inner core”. The word ‘elastic’ in the first line was erroneously replaced with ‘electric’. □

cAMP-induced switching in turning direction of nerve growth cones

Hong-jun Song, Guo-li Ming & Mu-ming Poo

Nature **388**, 275–279 (1997)

The order of panels in Fig. 3 of this Letter is incorrect as published. Figure 3a–e should be labelled as f–j, and Fig. 3f–j should be labelled a–e. □

corrections

Synthesis and X-ray structure of dumb-bell-shaped C₁₂₀

Guan-Wu Wang, Koichi Komatsu, Yasujiro Murata & Motoo Shiro

Nature **387**, 583–586 (1997)

In this Letter, we overlooked a citation of G. Oszlanyi *et al.*, *Phys. Rev. B* **54**, 11849 (1996), who reported the observation of covalently bound (C₆₀)₂²⁻ dianions from the X-ray powder diffraction patterns of the metastable phases of KC₆₀ and RbC₆₀. □

The complete genome sequence of the gastric pathogen *Helicobacter pylori*

Jean-F. Tomb, Owen White, Anthony R. Kerlavage, Rebecca A. Clayton, Granger G. Sutton, Robert D. Fleischmann, Karen A. Ketchum, Hans Peter Klenk, Steven Gill, Brian A. Dougherty, Karen Nelson, John Quackenbush, Lixin Zhou, Ewen F. Kirkness, Scott Peterson, Brendan Loftus, Delwood Richardson, Robert Dodson, Hanif G. Khalak, Anna Glodek, Keith McKenney, Lisa M. Fitzgerald, Norman Lee, Mark D. Adams, Erin K. Hickey, Douglas E. Berg, Jeanine D. Gocayne, Teresa R. Utterback, Jeremy D. Peterson, Jenny M. Kelley, Matthew D. Cotton, Janice M. Weidman, Claire Fujii, Cheryl Bowman, Larry Watthey, Erik Wallin, William S. Hayes, Mark Borodovsky, Peter D. Karp, Hamilton O. Smith, Claire M. Fraser & J. Craig Venter

Nature **388**, 539–547 (1997)

In this Article, we incorrectly stated that the amino acids lysine and arginine are twice as abundant in *H. pylori* proteins as they are in those of *Haemophilus influenzae* and *Escherichia coli*. This statement was derived from amino-acid analyses that compared absolute differences in abundance, but these do not reflect the frequencies with which amino acids are found in the organisms in question. The actual abundance of arginine in *H. pylori*, *H. influenzae* and *E. coli* is 3.5, 4.5 and 5.5%, respectively; the abundance of lysine in these organisms is 8.9, 6.3 and 4.4%, respectively. This oversight is particularly unfortunate because Russell H. Doolittle, who wrote an accompanying News and Views on our Article and brought this to our attention, was led to comment on the significance of our inaccurate observation. We regret this and any other misunderstanding that our error may have caused. □

Software solutions

This compilation covers a variety of chemistry packages for process monitoring, molecular visualization and kinetic simulation, as well as mapping and data processing software for geological and environment study.

ZooSeq

From Incyte

An animal gene sequence and expression database for drug development

This database focuses on genomic information for animals commonly used in pre-clinical studies of drug pharmacology and toxicology. It also contains bioanalysis tools for exploring the data and identifying genes of interest. ZooSeq allows scientists to compare gene expression, structure and function across species. For example, by correlating a drug's effects on a rat with the animal's genetic makeup, and then cross-referencing this data with LifeSeq, the company's human gene sequence and expression database, a researcher might better predict the drug's behaviour, efficacy and side effects before moving to human clinical trials. Each monthly data release for ZooSeq will contain DNA sequence and expression data for several whole-organ rat libraries. Initial sequencing efforts are focused on the Sprague-Dawley rat, but there are plans to expand the database to include tissue samples from other relevant animal models, such as mouse, dog and *Drosophila melanogaster*. The database will also target genes uniquely expressed in cells and tissues that are difficult to obtain in humans.

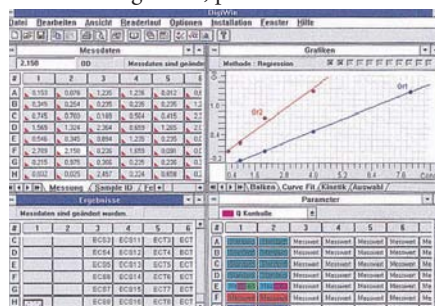
Reader Enquiry No. 100

DigiWin 3.12

From Asys Hitech

Windows-based ELISA microplate reading software

This program allows mouse control of each step, beginning with reader setup, plate layouts, identification numbers, all formulas for calculating results, including statistical functions and calibration curves, with formulas that can be defined in usual mathematical manner. Different calculation levels are possible and results of calculation levels can be combined and calculated, as well. Moreover, reader management, plate identification and



DigiWin Windows-based ELISA microplate reading software from Asys Hitech.

storage of data is controlled within the program. Each printout can be individually configured for customized output. Thus, there is documentation of reader values, sample identifications, calculation formulas and data presentation, such as bars or curves, with different algorithms. DigiWin can also manage kinetic readings, kinetic calculations and temperature control.

Reader Enquiry No. 101

Advanced Quantifier

From Bio Image

A new family of one-dimensional electrophoresis gel analysis systems for the PC and PowerMacintosh

The Advanced Quantifier (AQ) product line includes the AQ 1-D and the AQ 1-D Match systems. The systems allow for a choice of scanners and cameras, which are said to provide high resolution and a wide dynamic range to get the best data from gels. Using proprietary algorithms, AQ software will automatically find lanes and bands, and with simple commands quantify bands and calculate their sizes from standards giving accurate and reproducible results. AQ 1-D Match system enables the user to build a database and match over a thousand lanes, reporting the results as a phylogenetic tree, a contig map, a matrix or a table. All reports can be fully customized by the user and printed or exported for use with other packages. AQ software is also available separately for use with existing hardware, will accept any TIFF file and can drive any TWAIN compliant scanner, says Bio Image.

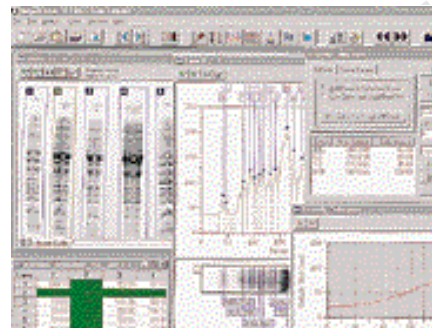
Reader Enquiry No. 102

ImageMaster 1D Prime

From Hoefer Pharmacia Biotech

An upgrade to version 2.01 of this one-dimensional electrophoresis image analysis and database software

With ImageMaster 1D Prime and Elite electrophoresis gel image analysis software designed specifically for the 32-bit environment of Windows95, both editions provide mathematical corrections that enable the user to extract reliable qualitative and quantitative information. This is true, says the manufacturer, even from gels that might commonly be thrown away. The software provides extensive user-specified reports to document and present raw data and results of the analysis. This addition to ImageMaster 1D Elite allows the archiving, query retrieval, sorting and comparison of lanes from a large number of images and experiments. It is well suited for



1D Prime 2.01 database software from Pharmacia.

population studies of DNA or protein marker frequencies, for analysis of historical electrophoretic quality, for recording of a manufacturing process and for identification, typing or classification of new organism isolates. The flexible reporting includes comparison tables with user-selectable fields and dendrograms to illustrate relatedness. The application software structure is presented in the same fashion as Windows95 Explorer for flexibility, ease-of-use and quick learning.

Reader Enquiry No. 103

Spectrophotometry

From ScienceMedia

A new multi-media tool that helps students perform a higher number of successful laboratory experiments

Developed by Barbara Sawrey and Gabriele Wienhausen at the University of California, San Diego, this product allows students to practice with a virtual working model of a spectrophotometer. This interactive unit also provides explorations and animations about the theory of light absorption, the relationship between energy and light at the atomic level, as well as Beers law. Students in chemistry, biochemistry, biology and molecular biology can work at their own pace to develop expertise in this fundamental piece of laboratory equipment.

Reader Enquiry No. 104

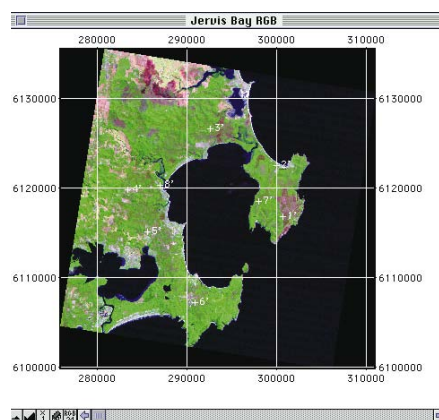
Mapping and geography software

Dimple 3.0

From Cherwell Scientific

Version 3.0 operates for the first time on Windows and Macintosh platforms

Dimple has been adopted in many areas of remote sensing, such as exploration, land and forestry management, and environmental protection and is able to process a wide range of data from satellite and airborne data to radar, digital terrain, geochemical or geological data. The new version also con-



Dimple 3.0 data processing software can work with airborne and satellite data.

tains new functionality to meet the demands of remote sensing analysis, including full processing capability for 32-bit, real- and floating-point data, direct colour processing and full annotation capability for map production. A free demonstration of Windows or Macintosh is available on CD-ROM and includes a fully functional copy of the program and sample data, except that the save and print functions are disabled.

Reader Enquiry No. 105

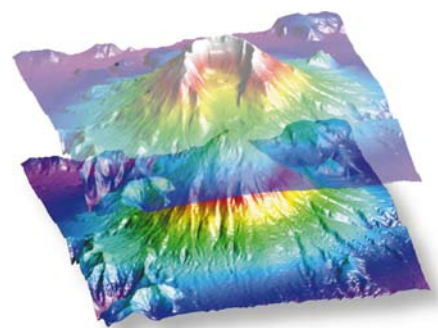
ER Mapper 5.5

From Earth Resource Mapping

Advanced three-dimensional graphics that features 'multi-surface 3D' under Windows NT, Windows95, Sun Solaris and SGI Irix

ER Mapper now has integrated three-dimensional viewing and printing for realistic viewing and easy interpretation. It can prepare and print photorealistic images, even on 'small' machines. It is capable of contour generation, enhanced map composition and Windows95-style wizards that help automate frequently used operations. Arc/Info-compatible files now support reading, writing and annotating of PC Arc/Info (dBASE) format coverages, as well as the Unix-style coverages. The program offers scripting language so that users can create their own 'wizards'. There is also single- or multi-colour contour generation from gridded data at user-specified spacing and line types.

Reader Enquiry No. 106



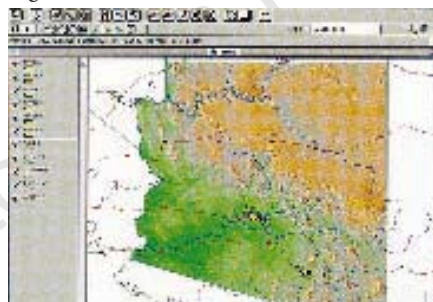
Three-dimensional mapping software from Earth Resources Mapping.

Digital topography maps for GIS

Chalk Butte

A set of full-colour, shaded relief maps that can be imported into and georeferenced in GIS software

For use with GIS software like ArcView and Arc/Info, these raster maps constitute a highly detailed topographic base over which vector data features, such as streams and roads can be overlaid. The map images have been computed from a 6-arc-second subset of the USGS 3-arc-second gridded, elevation database. Coverage includes the entire United States with the exception of Alaska (all files are uncompressed TIFFs). Each state has its own directory and within that directory are all of the one-by-one quadrangles of which the state is composed. There is also a screen-sized map that shows the entire state. The scale on these one-by-one degree maps is 1:530,000. Images include 24-bit colour, such that two or more map tiles may be assembled to create larger maps either in GIS or image processing software.



US digital topography for GIS from Chalk Butte.

Reader Enquiry No. 107

Chemistry collection

CAChe System

From Oxford Molecular

Version 3.0 of this experimental chemistry software for Windows95 and NT

The new release includes the ProjectLeader interface that lets chemists solve research problems by focusing on properties instead of complex computational methods. This interface, previously available only as a Macintosh version, is designed to make computational chemistry easily accessible to experimental chemists without requiring them to become computational specialists. It should help chemists to explain the activities of compounds, model reactions and predict a wide variety of chemical and physical properties for small molecules, including those containing metals. The company has announced a new lower-cost licensing option for the CAChe WorkSystem family, which includes a floating licence for concurrent users. A free trial programme is also offered, whereby current owners of any CAChe product who have current maintenance or bring their maintenance current, can try WorkSystem on either Macintosh or Windows for 90 days and then choose to upgrade, side-grade or keep both the new version and their existing versions at a reduced price.

nance or bring their maintenance current, can try WorkSystem on either Macintosh or Windows for 90 days and then choose to upgrade, side-grade or keep both the new version and their existing versions at a reduced price.

Reader Enquiry No. 108

Chem-X

From Chemical Design

Software for drug discovery is now available for use on Digital Equipment's Alpha processors running under Windows NT

This latest implementation extends the range of platforms on which Chem-X will run with the same user interface, whether in stand-alone or client-server environments. The Alpha NT version of Chem-X provides access to the entire range of software tools for exploiting chemical and biological information. Covering the various stages of the drug discovery cycle, Chem-X's computational capabilities are fully integrated with combinatorial chemistry tools, automatic programming of synthesis robots, reagent and plate-inventory management, as well as transparent links to biological testing results in Oracle. Chem-X for Alpha NT is currently supported for the Windows NT Workstation 3.51 operating system and later versions.

Reader Enquiry No. 109

ChemOffice 4.0

From CambridgeSoft

Simultaneous release of ChemOffice 4.0 and ChemDraw 4.0

The ChemOffice suite is an integrated desktop environment for chemists, featuring applications for structure drawing, molecular modelling and chemical information management. Both packages are available for Windows and Macintosh, as are the new Ultra versions of each, which include additional capability to enhance productivity in the chemistry laboratory. ChemDraw Pro 4.0 has a database query capability, chemical intelligence and publishing flexibility. It is designed to communicate structures and mechanisms faster and easier than before, both in print and over the Internet. Users can calculate such chemical information as molecular weight, empirical formulae and elemental composition. Moreover, there is a chemical syntax checker, a facility for structure clean-up, automatic conversion of atom labels to actual structures, and the ability to draw organometallics and isomer mixtures. ChemOffice integrates Chem3D Pro and ChemFinder Pro to provide a set of software tools that can be used to observe low-energy conformations with molecular mechanics, to do molecular dynamics simulations and to explore protein domains by

quickly identifying residues. ChemFinder is a database management system for chemists with a search engine — it allows users to access and search databases of chemical information, as well as to organize and cross-reference data. The Ultra packages offer Mopac Pro for semi-empirical modelling, ChemInfo Pro for instant access to current chemical information (including the Maybridge and Acros chemical catalogues, the US National Cancer Institute's archive of compounds, current data from ISI's Index Chemicus service, and more), and ClipArt Pro for easy one-step graphics and publishing design work.

Reader Enquiry No. 110

ACSL Optimize

From MGA

A new PC-based software solution that performs chemical kinetic simulations

Targeted at chemical industry research and development applications, ACSL Optimize utilizes a graphical interface to assist in performing parameter estimation, optimization and sensitivity analysis on numerous analytical chemistry applications from models of chemical reactors to toxicology and environmental simulations. Like MGA's other simulation products, this

product is based on ACSL (advanced continuous simulation language), commonly used for modelling nonlinear, continuous systems. This new software should help engineers design chemical processes that have higher yields, lower costs and more efficient use of resources. Toxicologists can use it to estimate model parameters that are difficult or impossible to create in the laboratory. ACSL Optimize's functionality is based on a 32-bit architecture that enables users to perform highly complex analysis, visualization and post-processing of large amounts of data quickly.

Reader Enquiry No. 111

ChemWeb.com

From MDL

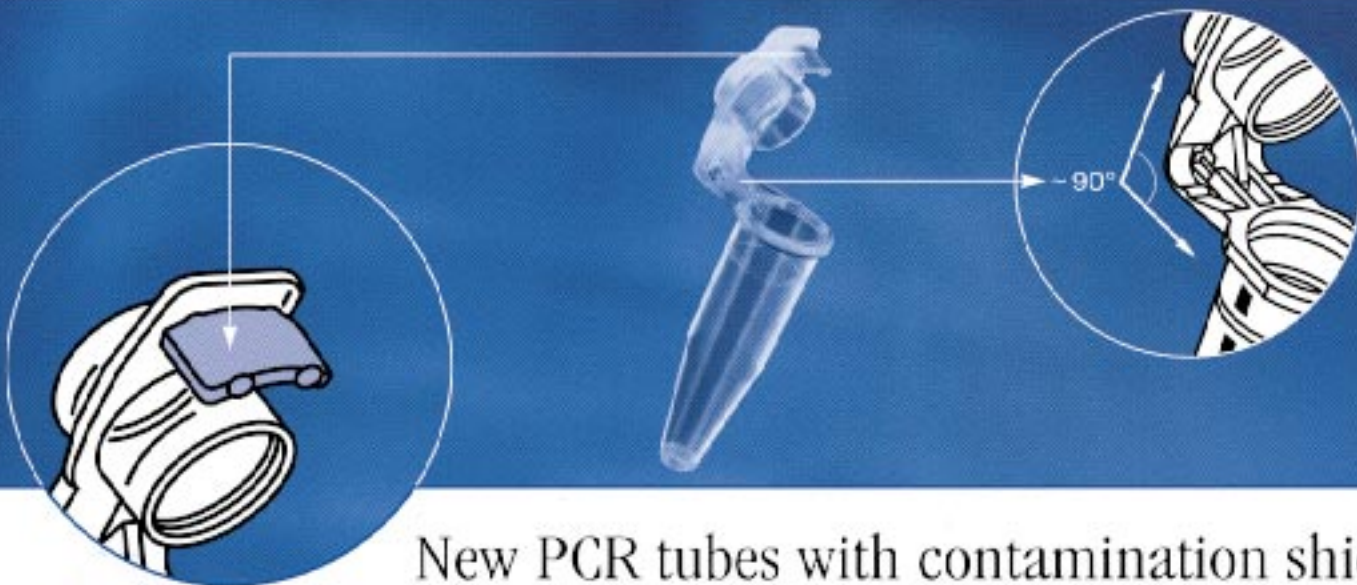
New additions to the website include databases, journals and online publishing for chemistry research <www.chemweb.com>

This site has undergone a redesign of its interface and improvements in technical performance and capabilities. Access to it enables researchers to search, retrieve and manipulate chemical information by structure. Also redesigned is the BiblioteK text search engine, which is stated to run many times faster. There are two database systems that streamline business and administrative

aspects of research: the Available Chemicals Directory (ACD) and the OHS safety series database. The former cover some 595,000 products from some 300 manufacturers and the latter database has almost 54,000 material safety data sheets (MSDSs) from MDL, covering pure substances and commonly used chemical mixtures. Furthermore, there is access to the Investigational Drugs database (IDdb), which was developed with the help of major pharmaceutical companies as a source of evaluated information on drugs in research and development at over 12,000 companies and institutes. It can combine results from searches of free-text, controlled-index terms and chemical structures. Two products derived from IDdb are available separately, as well: *Investigational Drugs Daily Highlights* covers breaking news of drug research from newswires, scientific meetings and journals, business sources and patents; and, *ID patent fast alert* provides a weekly update on pharmaceutical and biotechnology patents. The revised site will be cross-searchable with all other journals in the ChemWeb.com library and renders active chemical structures using MDL's Chemscape technology database of some 2,000 macromolecular structures, reported since 1991.

Reader Enquiry No. 112

ADVERTISEMENT



● **Accessories:** The work tray and the accompanying frame may be used as a tube rack or for inserting tubes into the thermoblock.



The innovative contamination shield on the lid of the new 0.2 ml PCR tube prevents accidental contact with the inner surface of the lid during opening and closing.

The tubes also have a special hinge which allows the lid to click into a set position of approx. 90° and thus prevents overlapping with adjacent tubes.

The new, thin-walled PCR tubes from Eppendorf are available as individual tubes, and also as strips of five or eight tubes.

New Eppendorf Thermocyclers:
Mastercycler® gradient and Mastercycler® personal.
Ask for your information pack now!

eppendorf

Eppendorf - Netheler - Hinz GmbH · 22331 Hamburg · Germany · Tel. +49 (40) 53601-0 · Fax +49 (40) 53601-556

e-mail: eppendorf@eppendorf.com · eppendorf home page: <http://www.eppendorf.com>

Maths, graphs and chemometrics

TableCurve 3D

From SPSS

Flexible surface fitting in Windows

Engineers and scientists can now fit and rank 36,000 out of more than 450 million frequently encountered equations in one automated step. Version 3.0 of this software completely automates the process of surface fitting and gives engineers and scientists assistance in the discovery of the a model that best fits their three-dimensional empirical data. There are thousands of equations to help solve complex science and engineering problems more efficiently. After fitting and ranking the best equations for the data, the program provides the users with all the graphical and numerical information they need to discover visually their preferred model. The program has an 'animation' feature that allows users to view the nuances within the fit. TableCurve 3D includes enhanced linear fitting with high-order, true bivariate models, which can be used to model complicated surfaces, even without knowing the nonlinear form of the surface. The user can also do non-parametric estimation of gridded and scattered data: this includes non-parametric fitting procedures and provides five different algorithms for gridded data and ten different algorithms for scattered data. As a result, users can now model complex data sets that can not be estimated using parametric models. The data manipulation and graphic enhancements include two smoothing operations that offer the intrinsic smoothing of NURBS (non-uniform rational B-splines) for gridded data and an enhanced three-dimensional Loess procedure for scattered data. It also can plot any of the five first or second partial derivatives.

Reader Enquiry No. 113

Modde 4.0 and Simca-P

From Umetri

Modde is a Windows-based package for statistical experimental design and analysis

It is said to be an easy-to-use, graphically oriented program with features that include a 'plot wizard' to help in the creation of two-, three- and four-dimensional contour plots and an interactive optimizer for the simultaneous optimization of an unlimited number of responses. Simca-P is also a Windows-based graphical package, but this program has been designed for multivariate data analysis and process modelling. Using multivariate control charts derived from PCA and PLS, it provides a complete picture of the process, taking account of all variables and responses simultaneously. Data is used to model the process and predict the down-

stream effects of any deviations. A related product, Simca 4000 On-Line, will acquire data from the process in real time to show in graphical format how well or poorly the process is performing from moment to moment.

Reader Enquiry No. 114

Mathematica

From Wolfram Research

Smarter new maths abilities and enhanced windows performance in version 3.0.1

This version of Mathematica provides speed increases in almost all kernel operations under both Windows95 and Windows NT, according to the manufacturer. For Pentium processor machines, many typical calculation-intensive operations, such as large matrix manipulations, are now stated to be as much as 50 per cent faster. It performs more algebraic transformations, sums, simplifications and symbolic integrations than before. The company has also added

Japanese language compatibility on the Windows and MacOS platforms. Japanese language support includes copy and paste, native Japanese filenames, printing to PostScript and HP-PCL printers, as well as the use of Japanese text in dialogue boxes, such as search and replace. There are language kits for German and French, as well (others are in preparation). This program can export notebooks in HTML format for presentation on the Internet. Wolfram states that on some types of graphics, the memory usage has been reduced by as much as 90 per cent. Colour gradations and other features of colour printers are now fully supported, and PostScript font handling is also said to be improved.

Reader Enquiry No. 115

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

x/STAT Now you can beef up the statistical capabilities of your Excel™ spreadsheet with the x/STAT add-in. Performs ANOVA/ANCOVA, multiple regression, discriminant and correspondence analyses, multidimensional scaling, clustering, goodness of fit tests and tabular sorting. Optional generalised fitting, Logit/Probit and Exact tests.

SIMSTAT Simstat for Windows' innovative output management features and scripting language lets you automate statistical analysis and write interactive tutorials. Performs a wide range of stats, including multiple regression, GLM ANOVA, time series, many non-parametric, bootstrapping.

Kovach Computing Services
Tel/Fax: +44-(0)1248-450414
info@kovcomp.co.uk, <http://www.kovcomp.co.uk/>

READER ENQUIRY NO. 16

Molecular Biology Certified™

Over 400 recipes and formulae in stock

CONVENIENCE
CONSISTENCY
ECONOMY
RELIABILITY
SELECTION

BIO 101 has the largest selection of Bacterial & Yeast Media available in capsule, powder & pouch form

Call us or your local distributor for ordering & information
[800] 424-6101 ♦ fax: (760) 598-0116
e-mail: info@bio101.com ♦ www.bio101.com

READER ENQUIRY NO. 15

CUSTOM HYBRIDOMA DEVELOPMENT

- Ascites Production
- Purification • Conjugation
- Polyclonal Antibodies

Also Available

1-800-481-9737

e-mail: antibodies@htbio.com
<http://www.htbio.com>
fax: (760) 788-9694
P.O. Box 1319
Ramona, CA 92065

HTI
HTI Bio-Products, Inc.

READER ENQUIRY NO. 96

Check out our homepage at <http://www.resgen.com>

Peptides
\$18/residue

Antibodies
monoclonal - \$8500
polyclonal - \$1195

Complete services for your custom peptide and polyclonal or monoclonal antibody needs.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 205-536-9016

READER ENQUIRY NO. 87

Bioinformatics in a post-genomics age

Career opportunities abound in this fast-expanding area of biological research. From computational biology to drug development, jobs are there for the taking. How long can the boom continue?

Diane Gershon

One only need flick through the classifieds in the scientific journals or scan company Web sites to see abundant job vacancies in bioinformatics. Many feel the subject has come of age in the past few years, to become a *bona fide* discipline.

As David Searls, director of bioinformatics at SmithKline Beecham (King of Prussia, Pennsylvania), points out, bioinformatics is supported by theory; an increasing number of journals and scientific meetings are devoted to it; and it now has its own society, the International Society for Computational Biology (associated with the conference series Intelligent Systems for Molecular Biology), whose president is Larry Hunter of the National Library of Medicine.

In the early 1990s, genomics changed from an essentially academic exercise to a serious commercial endeavour, says Chris Fields, vice-president and chief scientific officer of the bioinformatics start-up company Molecular Informatics (Santa Fe, New Mexico). High-throughput sequencing became feasible on an industrial scale, first being done by the Institute for Genomic Research (TIGR, Gaithersburg, Maryland), Human Genome Sciences (HGS, Rockville, Maryland) and Incyte Pharmaceuticals (Palo Alto, California), and then by many others, Fields says.

The 'commercialization' of genomics was repeated a few years later in bioinformatics, leading to the emergence of a number of companies. One of the initial challenges was to provide researchers with access to the vast amounts of biological information in major international public and proprietary databases held in various formats (there are attempts afoot to develop common standards linking biological databases).

Then there is the traditional challenge of sequence analysis, which, Searls says, "is pushing into what's called the 'twilight zone' of more and more distant similarity between homologous proteins — which is the first, and best, clue to function". There is also the issue of relating sequence to function via structure, for which structural biology is crucial.

The buzzword right now is 'post-genomics', says Searls — that is, what do you do once you have the sequences of all the human genes? "I think a lot of post-genomics will have to do with understanding networks, pathways and cascades, signal transduction, metabolism and genetic regulation. So there are all sorts of computational and mathematical issues to be resolved and to be brought to

bear on these problems," he says.

Although the term 'bioinformaticist' is used frequently, very different types of people end up working in the field. First there is what Fields dubs the "canonical bioinformatics person", someone who knows a lot of molecular biology and is an expert user of various types of mainly sequence-analysis techniques, and who can interpret the meaningfulness of the results obtained by such techniques. This is the sort of person, he says, "who can look at a marginal BLAST search and say 'yes, this is a real hit' or 'no it's not, it's noise'".

Then there are software developers, who know how to write code, how to build databases, and who are able to connect various bits of third-party software together to produce a system. People do cross over and become code writers, says Fields, but often they become light-duty code writers and are unlikely to have the experience required to write real software applications or to design and build database systems.

Also included in the mix are people who started as mathematicians and physicists who have reinvented themselves as bioinformaticists.

Hot market

There is little question that the bioinformatics job market is hot right now. Most companies — large and small — are actively recruiting in this area. The main issue for them is that there are not enough 'qualified' people.

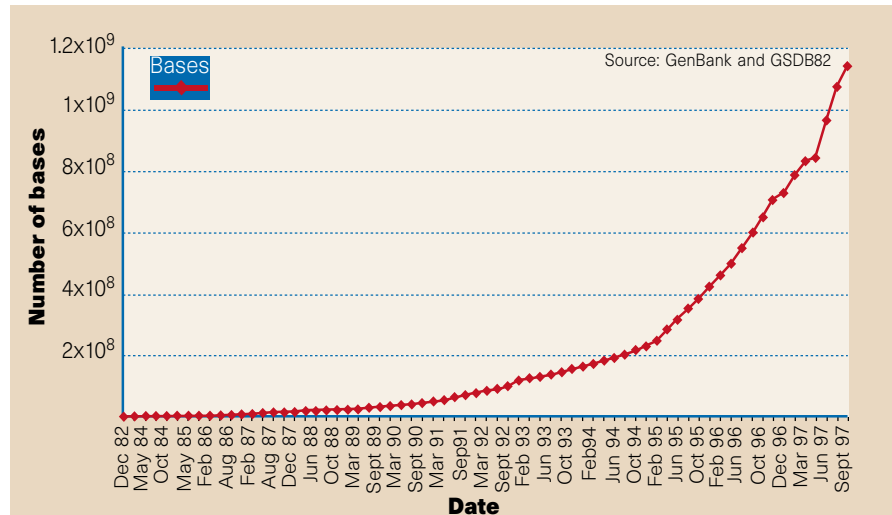
A case in point is Structural Bioinformatics (San Diego, California), a start-up company that, as its name suggests, is particularly interested in structural information about gene products. The company has been look-

ing for a vice-president of bioinformatics since December — someone who takes a systems approach to structure-function issues, has a strong grounding in biology, cell biology and biochemistry and who knows how to use computational systems to solve these problems, but who is not necessarily a computational scientist.

Founded in March 1996, the company currently has 15 full-time employees, 10 of whom are PhDs: for the most part biophysicists and computational scientists with a grounding in molecular modelling/molecular dynamics and computational technologies related to virtual screening/combinatorial screening.

"I've seen people hired right out of a doctorate programme, with no post-doc and no commercial experience," says Alan Engelberg, vice-president of marketing at NetGenics (Cleveland, Ohio). Such is the demand, he says, that he has heard of instances where bioinformaticists — PhDs with experience in molecular biology and with some knowledge of computational systems but without postdoctoral experience — command starting salaries up to \$90,000.

Fields says Molecular Informatics tends to look for people with some experience in the commercial sector and is "more interested in applicable experience than in academic credentials". Formed last year as a commercial spin-off of the National Center for Genome Resources (Santa Fe, New Mexico), Molecular Informatics now has a staff of 50 and already has two products on the market — BioMerge and and BioLIMS — for the management of genome data. The company is almost entirely focused on software devel-



Explosive growth: the rise and rise of public DNA sequences, compiled by Bruno Sobral at the National Center for Genomic Research

opment although, as Fields points out, that ranges from people who do requirement analysis and do not write software applications, to code writers, to people who do testing and documentation. The ability to function well as a member of a team is critical to software engineering, not just as a skill but as an attitude, says Fields.

There are signs investors are beginning to take notice of these bioinformatics start-ups. Both NetGenics and Structural Bioinformatics have agreed financing deals with venture-capital companies. And what these companies lack in terms of the resources and infrastructure of a large pharmaceutical company, they often make up for by a lively and enterprising work environment. The fact that everybody gets a piece of the action through stock options can also be a draw. Although the world of large pharmaceutical companies is comfortable, it is sometimes "hard to move things quickly and in new directions", says Susan K. Burgess, vice-president of corporate development at Structural Bioinformatics and formerly with Wellcome and Glaxo.

But job seekers would also do well to take a look at what is happening inside some of the pharmaceutical companies.

Although generally slow to move into this area, SmithKline was one of the first companies to recognize the importance of integrating genomic data into the drug-discovery process. In 1993, it entered a \$125 million alliance with HGS, which gave it access to the company's vast database of human gene sequences. Searls says the link forced SmithKline into bioinformatics ahead of others and may have been one of the best side-effects of the HGS deal.

Indeed, it was the sheer scale of the bioinformatics enterprise at SmithKline and the company's commitment to academic-style research that lured Searls away from the University of Pennsylvania. SmithKline's bioinformatics group, mainly US-based, now numbers in the high 50s. In addition, the company has recently established a sizeable gene informatics group in the United Kingdom under the direction of Peter Goodfellow, consisting mainly of PhD-level biologists whose primary function is to discover new genes and genetic markers in the databases.

Searls currently has at least six job openings and is keen to recruit a number of senior researchers who he says would typically be academic types capable of running independent research activities.

Bioinformatics is clearly an area where supply has not yet outstripped demand and where people with the right skills can do very well — whether in drug or biotechnology companies or, now, in this new set of 'vendor' companies servicing the pharmaceutical and biotechnology industries. □

Diane Gershon is assistant editor, new technology, of Nature Medicine.

Common language of bioinformatics

Bruno W. S. Sobral



Bruno Sobral

There's no denying that bioinformatics is an exploding field. And as team leader for agricultural and environmental genomics at the National Center for Genome Resources (NCGR), a non-profit bioinformatics organization, I am struck by the lack of qualified people to meet the challenge.

The most effective solution for overcoming this limitation, as vast quantities of genomic data stream in from private industry and the public sector, are partnerships between groups specializing in bioinformatics and those in genetic mapping, physical mapping and genomic sequencing. But before going further, it is necessary to define some terms in the quest for a common language.

With the development of molecular biology over the past couple of decades, ever more powerful tools have been developed to dissect the role of genes in a wide assortment of traits, such as diseases of various organisms. In addition, new information has helped in the understanding of inherited traits and their influence on behaviour. More recently, as a result of computational and engineering advances engendered by the Human Genome Project, the field of genomics has emerged. Genomics has two major components: DNA sequencing in the molecular genetics laboratory; and bioinformatics, another new term which was spawned by genomics. Thus, genomics is a combination of complete nucleotide sequences for any organism and the information tools required for analysis of data. Bioinformatics is defined as the computational systems used to collect, store and analyse biological information. These include software systems that take in DNA sequence data, database systems that store data, and software systems that analyse stored data.

One of the key challenges in bioinformatics as a field is to move smoothly through the transitional period of collections of incompatible components to integrated systems. In the next few years, people working in bioinformatics will be tackling issues of interface homogeneity, development of scaleable software toolkits for data analysis, inter-database connectivity, consistency of terminology and long-term funding of public repositories of information.

Universities have not generally played a major part in bioinformatics because they have not had the resources. As genome pro-

jects broaden their scope from the human genome to model systems and agricultural and food genomes, universities need effective bioinformatics support. One solution is to enter partnerships with groups already experienced in bioinformatics, especially within the public sector. In this manner, DNA sequencing/physical mapping groups at universities can quickly develop large datasets without having to create their own bioinformatics infrastructure. NCGR, for example, has recently established a partnership with New Mexico State University to develop a national biotechnology information facility (NBIF — <http://nbif.org/>), a five-year, \$8.5-million partnership. NBIF is developing a plant-specific metabolic pathways database while providing training and bioinformatics support for graduate students and postdoctoral fellows. NCGR's strategy is to provide similar bioinformatics support to leading agricultural research institutions.

Through such partnerships, collaborative development of human resources in bioinformatics can become reality. Not only will this provide much-needed bioinformatics expertise in universities; it will also provide the private sector with individuals trained to make use of genomic information. As genetics and biology enter the twenty-first century, it is clear that genomics is changing the way biological research is done.

Public information and biological reagent repositories are a decentralizing and democratizing force in research, much as the Internet was and is for computers. Eventually, any scientist may have direct and rapid access to information and reagents without needing to create and maintain a large-scale operation. I hope that this will also mean that scientists will be rewarded more for creativity and less for their fund-raising capabilities. □

Bruno W. S. Sobral is at the Genomics National Center for Genome Resources (NCGR), Santa Fe, New Mexico 87505, USA. e-mail: bws@ncgr.org.

Crossover in interest and skill

Brendan Horton

Barry Robson and Roger Pain wrote one of the first papers discussing how to store protein sequences and recover information (*Nature* 227, 62–63; 1970). As he has done throughout his career, Robson prefers to erase the artificial lines that separate and define disciplines, describing himself on his *curriculum vitae* as "an experimental medical scientist and a computational chemical physicist". This mixture of 'hard' **Barry Robson**



and 'soft' science gives him much enjoyment in his work.

Because of mild dyslexia, Robson was channelled away from maths and theoretical sciences by the age of 16 in the UK education system. Unrecognized in his youth, Robson's dyslexia affected his ability to spell, make left and right spatial judgements and do arithmetic. In particular, he notes that at that time it was not appreciated that mathematics and programming are not the same as arithmetic.

It was not until after Robson had been trained as an experimental medical scientist that he discovered he could do theoretical science after all. In his experience, students were channelled into subjects depending on their early academic performance, as if the proper intellectual hierarchy dictated mathematics for the most able, followed by physics, chemistry, biology, psychology and finally sociology. Robson's first self-directed step to theoretical biology came when, at 18, he developed mathematical functions for modelling molecules in three dimensions.

"A lot of protein modelling code sprang from those early routines, which were first implemented to draw molecules with the help of a mechanical calculator," he says. Realizing that the mathematics would save him work, he began "to collect maths tricks like a magpie collects shiny objects". He remains, however, envious of those who have been formally trained and have a full repertoire of mathematics skills.

Yet it was the lack of formal training that benefited his work in a series of information theory papers with Pain, Eicichi Suzuki and others from 1970 to 1976, leading to the Garner-Osguthorpe-Robson (GOR) secondary-structure prediction method. To deal with very sparse X-ray data from 1968–1972, Robson used bayesian statistical methods, which were largely ignored at the time.

"There was in fact a very bad reaction to anyone using such techniques, and except in very rare and very advanced courses, they were not taught at all." They were regarded as 'subjectivist', he says: "I would never have chosen that worthy approach if I had been formally trained in the approved statistical philosophy." Today, he says, bayesian methods are the mainstream of bioinformatics.

It was these early successes, perhaps, that strengthened Robson's belief that lines drawn between fields were artificial and limiting. The 1980s were revolutionary, bringing the ability to engineer proteins through genetic engineering; large increases in requisite data by breakthroughs in NMR, crystallography and monoclonal antibodies; and supercomputers to run these programs, as well as cheap, networked computing to process and share data locally.

Robson's involvement in commercial and semi-commercial ventures led to the formation of the Epsitron peptide and protein engineering research unit at Manchester

University, and the Proteus group of pharmaceutical companies. He stayed as reader in biochemistry at Manchester while cofounding and becoming scientific director of Epsitron and Proteus. This year he became a visiting researcher at Stanford, as well as a principal scientist at MDL, where he helps to acquire and develop new tools, trains staff and customers, assists marketing and sales, and advises on general strategy with MDL's Bioinformatics Workbench and Molecular Informatics' BioMerge systems.

Robson participates in a course at Stanford which is open to those from other departments, including medical students. He says that programmes like Stanford's are mainly concerned with preparing the next generation of bioinformaticists and tool providers through research (see Internet table). Robson is unaware of any example where someone from industry can take a short course in an academic institution and then return to industry, but notes that the master's programme at the University of Manchester is

mainly populated by industry researchers, indicating that there is such a need.

There should be rich opportunities in this field, but when will it reach saturation? Will all the hyperbole result in the overproduction of bioinformaticists similar to the glut of molecular biologists that resulted from unchecked growth in that field in the 1980s and 1990s? The expertise shortage is a real bottleneck, but there are ways in which computing and artificial intelligence (AI) can help, says Robson. These include building problem-solving environments to make best use of a limited number of experts; automating these to capture expertise (moving towards 'expert systems'); adapting them to smarter, more human-independent AI systems, still further reducing the dependence on experts; and adapting 'expert' systems as teaching machines.

Ultimately, advances of this kind will limit and eliminate jobs. The National Academy of Sciences' office of scientific and engineering personnel met earlier this month to examine

Table 1 Internet resources for bioinformatics

Organization	http://
Affymetrix	www.affymetrix.com
Human Genome Sciences	careers.hgsi.com
Incyte Pharmaceuticals	www.incyte.com
Molecular Informatics	www.molinfo.com
MDL	www.mdli.com
Molecular Applications Group	www.mag.com
National Center for Biotechnology Information	www.ncbi.nlm.nih.gov
National Center for Genome Resources	www.ncgr.org
National Center for Supercomputing Applications	www.ncsa.uiuc.edu/Apps/CB
NetGenics	www.netgenics.com
Merck	www.merck.com
Pangea Systems	www.PangeaSystems.com/0_home/index.htm
Pfizer	www.pfizer.com
SmithKline Beecham	www.bioinformatics.com
Structural Bioinformatics	www.strubix.com
The Institute for Genomic Research	www.tigr.org
Stanford Groups	
Sam Karlin Group (mathematics)	genomic.stanford.edu
Michael Levitt (structural biology)	hyper.stanford.edu/faculty.html
Section on Molecular Informatics – (acad. progs)	www-SMI.Stanford.EDU/smi/academics
Douglas Brutlag (bioinformatics group)	motif.stanford.edu
Section on Molecular Informatics	camis.stanford.edu
Program in Molecular and Genetic Medicine	mgm.stanford.edu/br/
General	
Weizmann Institute Genome and Bioinformatics	bioinfo.weizmann.ac.il/
Bioinformatics Services	www.bioinformatics-services.com/
Bioinformatics at University of Wisconsin	www.bocklabs.wisc.edu/Home.html
Bioinform	www.bioinform.com/
Biocomputing	www.cryst.bbk.ac.uk/BCD/bcdgloss.html
BioComputing Hypertext Coursebook	merlin.mbcrc.bcm.tmc.edu:8001/bcd/Curric/welcome.html
Genomics	www.phrma.org/genomics/index.html
KEGG Encyclopedia of Genes and Genomes	ep3000.scl.genome.ad.jp/kegg
The OMS Biology Resource	www.unl.edu/stc-95/ResTools/BioTools/biotools4.html
The DOE Biology Information Center	www.er.doe.gov/production/other/bioinfo_center.html
United States Department of Agriculture (Budget)	www.usda.gov/agency/obpa/Home-Page/obpa.html
Agricultural/Biotechnology Resources in Education	
Colorado State University, Department of Entomology	www.colostate.edu/Depts/Entomology/ent.htm
University of Arizona, Department of Entomology	ag.arizona.edu/ENTO/entohome.html
Iowa State University, Department of Agriculture	www.ag.iastate.edu/
University of Maryland, Center for Agricultural Biotechnology	www.umbi.umd.edu/~cab/index.html

methodological problems related to forecasting supply and demand for scientists and engineers and the use of such forecasts in policy making, as well to finalize the agenda for a workshop on these issues to be held early next year. According to the NAS, "Conflicting assessments that have emerged from recent analytical efforts have resulted in a considerable amount of confusion about the future state of labor market conditions for scientists and engineers." □

Brendan Horton is in Nature's Washington Office.

Choices and challenges

Potter Wickware

Computers have changed biology forever, even if most biologists don't yet realize it, says Michael Levitt, a structural biologist at Stanford University and the founder of Molecular Applications Group (MAG), in Palo Alto, California. Already, drug discovery is driven by the need to apply powerful computers to voluminous data sets, and the trend, he says, is certain to extend into all other disciplines in biology.

Chris Lee, Levitt's former graduate student and co-founder of MAG, agrees, noting that most biologists today use computers only in the most elementary way as a typewriter and graph-paper substitute. "Bioinformatics is really going to surge when biologists realize that there's a lot of value, and a lot of new insights, in being able to work across large amounts of data that they and all the other scientists in the world have produced," says Lee.

Levitt began using computers to solve problems in protein folding when working under John Kendrew, Max Perutz, Francis Crick and other eminent molecular biologists in the "golden age" of the 1960s at the Laboratory for Molecular Biology at Cambridge, United Kingdom. Today the 7lb laptop he carries in his backpack has more than a thousand times the computing power, at less than a thousandth of the cost, of the punch-card behemoths of 30 years ago.

Accompanying the relentless increase in computing power is a breathtaking expansion of biological data from the human and other organism genome-sequencing projects. Complementary information from the pharmaceutical chemistry, neuroscience, microbiology, immunology, clinical trials, toxicology, teratology, epidemiology and other disciplines waits to be integrated with the genetic and structural data. There is no way to obtain a global view of all this information, to establish links between disparate fields of knowledge, without the computer.

Myra Williams, MAG's new president, has a PhD in biophysics from Yale and was hired this summer from Glaxo Wellcome to

launch the company's GeneMine Pro suite of bioinformatics tools. She observes that rates of data acquisition, far from levelling off, are accelerating. Soon innovations like Affymetrix's high-density oligonucleotide array microchip will come online, generating terabytes ("terror-bytes") of new sequence information. How scientists navigate this ocean of biological information will be crucial.

"To be effective," Williams says, "bioinformatics tools must not merely automate data retrieval, but give researchers the information in usable form, through clustering, filtering, analysis and visualization, allowing them to perceive insights which might well have eluded them had they attempted to process the information manually."

The bioinformatics capabilities of MAG's GeneMine Pro program are built around its Discovery Engine, an automated Web browser which retrieves biological information in 22 categories from servers worldwide. After processing, the information is presented at the user interface, where it can be visualized in the context of sequence alignments and three-dimensional protein structures, or read as text. Large pharmaceutical companies, challenged by expiring patents and the high cost and slow pace of conventional drug development, are now the main source of sustenance for bioinformatics. The fact that genomics and bioinformatics are creatures of big industrial research environments inevitably leads to a blurring of distinctions between academic and industrial science.

Despite a strong and growing demand for bioinformaticists, there are few established training centres, perhaps 20 in the world, estimates Levitt. The field is still defining itself, and those who do have formal training are quickly snapped up by industry on hefty pay scales, leaving a deficit in the numbers of those available to train the next generation.

Nevertheless, Lee believes that the candidate who, on his or her own initiative, "can demonstrate the ability to cross over and generate results — not even necessarily original ones — is the one who will capture the recruiter's attention."

Levitt adds that the shortage of formal training slots coincides with exceptional opportunities for self-learning: with an Internet connection and inexpensive computer, one can download all the databases, programs and papers needed to undertake an original research project. "Spend two days looking at the results and thinking about what you don't understand. Use e-mail to contact someone who does, and ask, 'what should I be doing?'" he recommends. Prize-winning discoveries can be made in this way. "The problems are so difficult, and there is so much to be analysed, that the boom will not go away. It's a great time to be getting in. There's a wonderful lightness to the field." □

Potter Wickware is a science writer in Oakland, California, USA. e-mail: wick@netcom.com

Running to catch up in Europe

Helen Gavaghan

Across Europe, the story is the same. Demand for those skilled in bioinformatics exceeds supply. Like biochemistry and biophysics before it, bioinformatics is crushing the barriers between traditional academic fields, and demanding flexibility and a new way of thinking from its adherents.

Computational biology has meant different things to different people. Not too long ago, says Hans Prydz of the University of Oslo's Biotechnology Centre, it meant handling NMR data or analysing Doppler echograms. Now renamed bioinformatics, it means looking for patterns in DNA and RNA, predicting protein structure, modelling proteins and mining massive databases that continue to grow. When the DNA database run by the European Bioinformatics Institute (EBI) was first set up, it contained 700,000 nucleotides: now there are more than a billion.

Driven by the scientific and commercial importance of bioinformatics in genomics and drug discovery and development, governments, universities and industry are responding with varying degrees of vigour and success to the skills shortage and are seeking ways to cross the boundaries between disciplines as diverse as engineering, physics, mathematics, computer science, statistics, protein chemistry, genetics and molecular biology.

At European level, the EBI, based near Cambridge (United Kingdom), is funded to the sum of about DM 9 million (\$5 million) by members of the European Union and Israel via their contributions to the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany. Contributions from the pharmaceutical and agricultural industries roughly double the institute's income. The EBI, an offshoot of EMBL, develops tools for bioinformatics, seeks innovative ways to apply the tools, and runs training courses for academics and industrialists. Initiatives with industry include the Industry Affiliates Initiative, which helps small and medium-sized companies to identify and apply new techniques; the BioTitan Project, running nodes to enable faster access to databases; and the Biostandards project, funded jointly by industry and the European Union for promoting and developing standards.

National initiatives also exist, particularly in the United Kingdom and Germany. Says Andrew Lyall, responsible for bioinformatics at Glaxo Wellcome, "I think the UK is in pretty good shape." There are two government-financed initiatives in the United

Kingdom, both of which received a second lease of life earlier this month.

One of these schemes, supported by the Biotechnology and Biological Sciences Research Council (BBSRC), coordinates the UK bioinformatics community. Since 1994, the scheme has concentrated on developing software that would enable biologists without information technology (IT) skills to use some of the many tools important to their trade that are found on the World Wide Web. At a meeting earlier this month, the steering committee of the scheme decided to change the focus to training. Andy Brass, who runs a masters' degree course in bioinformatics at the University of Manchester and is a member of the committee, says, "We have a massive skills shortage. We need training at every level." The main aim will be to "train the trainers", says Brass, adding, "There are very few academics with the knowledge to teach bioinformatics because they have all been nicked by the companies." The view is echoed by others, and could account for Lyall's optimism.

The United Kingdom's second public scheme to promote bioinformatics is jointly funded by the BBSRC and by the Engineering and Physical Sciences Research Council. It too was started in 1994 (spending £1 million per year and making 25 awards in three years). This scheme was also renewed for a further three years earlier this month. Chris Miles, who heads the initiative for the BBSRC, explains that research proposals must have both a strong biological and computational component. During the next phase, researchers will address priorities such as finding methods for condensing and representing data visually and seeking informatics techniques allowing the prediction of protein structure from sequence data.

In Germany (see below) the country's main grant-giving body has launched a six-year bioinformatics programme, while in France, CNRS has allocated FF_r 15 million (\$2.6 million) for bioinformatics research this next year. Not all French scientists are convinced that they will see the money. Jean Thierry-Mieg from the CNRS in Montpellier

says that during the past five years spending in bioinformatics has been announced more than once, but because of political disagreements no money was forthcoming. The tide does seem to be turning, though. France is setting up a new sequencing laboratory in Paris and the University of Versailles is teaching a dedicated course on bioinformatics.

The picture elsewhere is less focused. Cecilia Saccone, a professor in the department of biochemistry and molecular biology at the University of Bari and one of Italy's leading bioinformaticists, is working with colleagues to design a bioinformatics doctoral programme for the country. Saccone has already started a project in bioinformatics, but Italy's national research council (CNR), although admitting that the subject is important, says that there is no money to support it.

Spain's leading light in bioinformatics, Alfonso Valencia from the Centro Nacional de Biotecnología in Madrid, is coordinating a group of researchers to promote training. Uniquely for a scientist, Valencia says: "We probably have enough money for bioinfor-

Germany on the trail of the Americans

Germany will not quickly end North America's lead in bioinformatics. But in the current biotechnology boom, the prospects seem bright.

Determined to make up for its lag in biotechnology research, a consequence of vociferous public opposition to genetic engineering, the German ministry of research (BMBF) has in the past few years launched several campaigns. Bioinformatics is a vital part of this initiative. Between 1993 and 1996, for example, the BMBF allocated DM37 million (US\$20 million) in research grants and pumped a further DM30 million into research institutes with a strong interest in bioinformatics. The BMBF plans to increase this support.

In addition, the Deutsche Forschungsgemeinschaft (DFG), Germany's main grant-giving body, is about to launch a six-year bioinformatics programme that will distribute about DM5 million in project grants in its first two years. The BMBF also has a scheme called Bioregio, a regional competition for a prize of DM150 million. Regions have

to present a collaborative programme involving industry and academic research institutes.

One company to benefit from the local contacts this programme promotes is Heidelberg-based LION Bioscience. Founded in spring this year, LION offers commercial services in bioinformatics and has ambitious plans to become Europe's leader. As a result of close collaboration with the European Molecular Biology Laboratory (EMBL), located on the same campus, it has been able to adapt EMBL's developments – efficient software and the fastest sequencer in the world – to its own needs.

LION is a sign of the good times that are almost certain to come for bioinformatics in Germany. The company started with 10 employees, will recruit a further 20 before the end of the year, and plans further expansion. "The technology was there, the human resources were there and we were just the first ones who had the idea to use it commercially. We filled a vacuum," says Peter Wiesner, senior manager of

business development.

The traditional large German pharmaceutical companies also have a growing need for specialists in bioinformatics, and find there are simply too few around. "Job opportunities in the field are so good that, at the moment, potential employers are finding the cupboard bare," says Hugo Kubinyi, head of drug design at BASF.

Academic institutes have even more ground for concern. "The battle between industry and academic research for competent bioinformaticists is, understandably, almost always won by industry," says Sandor Suhai, head of the department of molecular biophysics at the Deutsches Krebsforschungszentrum (DKFZ, the German national cancer research institute) in Heidelberg. In his group of 30 scientists (mainly graduate students) only two plan a career in academic research, says Suhai: industry can offer long-term employment, high salaries and sometimes more exciting jobs. He complains about the high turnover in his group and

says that he gets too few responses to advertisements.

The first university course for scientific informatics was established in response to this need in 1989 at Bielefeld. "We teach the basics of informatics but then focus on methods relevant to biology. The biology we teach is similarly restricted to that relevant to bioinformatics," says Robert Giegerich, dean of the technical faculty. Despite sceptics who believe the graduates are neither informaticists nor biologists, he has no doubt that his students will succeed on the job market. He believes bioinformatics will become a 'proper' science just as informatics separated from mathematics.

First employers' reactions to graduates from Bielefeld are enthusiastic. "In contrast to trained bioinformaticists, we have to teach an informaticist not only biology but also the basics of bioinformatics. This can take up to one year," says Martin Vingron, head of theoretical bioinformatics at the DKFZ.

Matthias Strobl

Matthias Strobl is a science writer in Munich, Germany

matics." But he adds: "We need more courses at all levels." The only course dedicated to bioinformatics in Spain is postdoctoral. Valencia is seeking formal relationships with the EBI as he believes that bioinformatics is not really promoted as a separate discipline in Spain. As in Italy, the money goes to molecular biology or protein chemistry groups that include a person with bioinformatics skills.

Away from academic institutions, the picture is equally vibrant and varied. Europe's pharmaceutical giants latched on to the significance of bioinformatic skills early on, and have indulged in aggressive poaching from the lower-paid public sector. The pharmaceutical industry's competition when recruiting high-quality staff, says Chris Rawlings, UK director for bioinformatics at SmithKline Beecham, is not from academic institutions, but from smaller genomics companies that can offer stock options and a prominent, leadership role in a small group.

Rawlings is less concerned by the competition for new recruits from the small bioinformatics companies that have recently begun the search for capital. The big pharmaceutical companies are building up a strong body of in-house expertise. "We go and talk to these people [in bioinformatics companies] and find out who they are. We are looking for niche technologies," he says. For the brave, the well informed and, above all, those with business acumen, there is nevertheless a whiff of money to be made from a start-up bioinformatics company.

"A lot of people see bioinformatics as the current technology frontier and believe, as they did eight years ago about the Internet, that there are a lot of opportunities out there," says Joseph Bergen of the venture-capital company 3i. But, says Bergen, it's all about timing, and bioinformatics has not broken cover yet. So far, he has seen interesting technology, but not the commercial sense that would convert an idea into hard cash. When the small companies do break cover, the SKBs and Glaxo Wellcomes of the world will be waiting. So far, says Lyall of Glaxo Wellcome, it is all smoke and mirrors.

For those contemplating a slightly less risky life in the commercial world, the big companies are still seeking people who will be effective from day one. "Bioinformatics is strategically important to us," says Rawlings. "The company's philosophy is that all new drug targets are expected to be discovered by genomics." SKB has two main informatics groups: in Philadelphia and Harlow (United Kingdom), whereas Glaxo Wellcome has three main centres for bioinformatics: Stevenage (United Kingdom), employing about 30 specialists; North Carolina, 25–30 specialists; and Geneva, between 5 and 10. There is also a small group in Madrid that collaborates with Valencia. Both Wellcome and SKB are not too far from Hinxton, where the EBI, the Sanger Centre (sequencing about a third

Useful Web sites for bioinformatics in Europe

The most useful site for those curious about bioinformatics in Europe is that of the EBI (<http://www.ebi.ac.uk>). The site is extensive, with a newsletter, job vacancies and a lot of educational material as well as further links, including one to its parent organization EMBL (<http://www.embl-heidelberg.de/>). It may just be my software, but I cannot always get through to this site.

A good starting point for a tour of European national centres is the

page of the European Science Foundation, which provides links to members. For example, (<http://www.esf-strasbourg.fr/members/map.htm>) will take you to French members, including the CNRS (<http://www.cnrs.fr/>). There is an English version of this site. One can also reach UK sites, including that for the BBSRC (<http://www.bbsrc.ac.uk/>). Finding information about bioinformatics on these sites requires some searching.

Also at European level is a page from the science directorate (<http://europa.eu.int/en/comm/dg12/biotech/bio-t-n.htm>). This provides links to biotechnology sites, including the European Federation of Biotechnology.

In the United Kingdom, the Wellcome centre provides a list of job vacancies at (<http://wisdom.wellcome.ac.uk/>). The page gives the option of searching for jobs in a given area, but I got no results. **Helen Gavaghan**

of the human genome) and the Medical Research Council's human genome mapping programme resource centre are sited.

SKB divides its bioinformatics activities into three areas: research; bioinformatics tool development and databases; and user support and services. This split accurately reflects the kind of jobs available to people interested in bioinformatics. The dream *curriculum vitae* for someone wanting to go into research in SKB or Glaxo Wellcome would show a bachelors' degree in a biological science, a masters' in computational science and a doctorate focusing on a problem in computational biology. Rawlings adds that he would like to see a stint or two as a post-doc. With this background, you are practically guaranteed to have Glaxo Wellcome and SmithKline Beecham fighting for your services. Says Brass: "You would probably earn in the mid to high thirties [£30,000+ a year], which compares with a salary in the mid thirties for a full professor in the UK."

The people working in bioinformatics tool development and user support and services will need a strong computing background. Of the support group, Rawlings says, "These people are not sitting on a help desk. They are performing sophisticated bioinformatics analyses for scientists without a Unix workstation." The pay scale, roughly — and both companies are a little coy about this — is low-to-mid £20,000 range for a person with a masters' degree, and high £20,000 to low £30,000 range for one with a masters' and a doctorate. "There is definitely a premium for having a masters' in bioinformatics compared with molecular biology," says Brass.

Given the boundaries that someone in bioinformatics must cross, perhaps the premium is justified. "It is not a field of science like genetics," says Michael Ashburner,

director of research at the EBI and a professor of genetics at the University of Cambridge, "it is a field in which nearly everyone comes from somewhere else." Slowly those people, in the United Kingdom at least, are being channelled through specialized masters' programmes. The universities of York, Manchester, Aberdeen and Birkbeck College (part of the University of London) offer masters' degrees in bioinformatics. There are more masters' courses in the pipeline, says Lyall, Edinburgh being one example.

Ashburner says: "We need to put more resources into training at the masters' level." The recent introduction of graduate schools into the United Kingdom on the US model (rather than having students attached to individuals in departments) makes the introduction of masters' degrees in interdisciplinary subjects like bioinformatics easier than it is in Europe, says Brass. The Manchester course takes up to 20 people each year; most come from a biology background, but Brass would like more from computer science, physics or mathematics.

The Manchester course concentrates on problem solving, being project- rather than lecture-based. The field is moving so fast that it would be difficult to construct a course otherwise. For example, when the EBI ran a workshop on 'common object request broker architecture' (CORBA — a way of transferring blocks of code between applications) earlier this year, it anticipated 40–50 attendees. In the event, 200 registered and a further 50 unregistered scientists turned up. Yet it is only this coming year that Manchester will teach CORBA. Even as they start their first job fresh from university, last year's graduates will find themselves running to catch up with the technology. □

Helen Gavaghan is a science and technology writer based in Hebden Bridge, Yorkshire, UK.